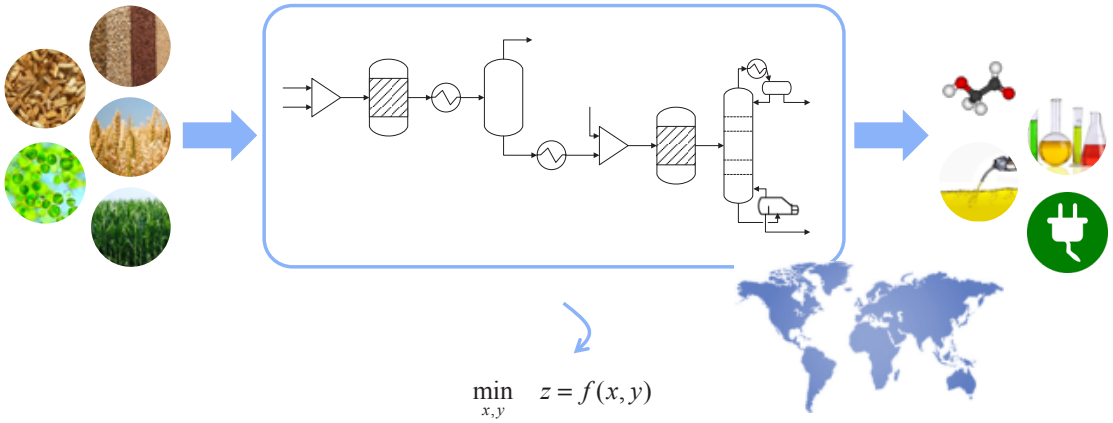
Modelling, synthesis and analysis of biorefinery networks

Bertran, Maria-Ona

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Modelling, synthesis and analysis of biorefinery networks



$$\min_{x,y} \quad z = f(x,y)$$
$$\text{s.t.} \quad g(x,y) = 0$$
$$h(x,y) \le 0$$
$$x \in R, \qquad y \in \{0,1\}$$

Maria-Ona Bertran
November 2017

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

PhD Thesis
Doctor of Philosophy

Modelling, synthesis and analysis of biorefinery networks

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Maria-Ona Bertran

November 2017

DTU

Modelling, synthesis and analysis of biorefinery networks

Maria-Ona Bertran

PhD Thesis, November 2017

Supervisors: Professor Rafiqul Gani, Technical University of Denmark
Professor John M. Woodley, Technical University of Denmark
Professor Anker D. Jensen, Technical University of Denmark

Examiners: Associate Professor Ioannis V. Skiadas, Technical University of Denmark
Professor Stratos N. Pistikopoulos, Texas A&M University, USA
Professor Jay H. Lee, Korea Advanced Institute of Science and Technology (KAIST), South Korea

Address: DTU Chemical Engineering
Department of Chemical and Biochemical Engineering
Technical University of Denmark
Soltofts Plads, building 229
2800 Kgs. Lyngby
Denmark

Phone: +45 4525 2800
Email: kt@kt.dtu.dk
Homepage: www.kt.dtu.dk

Modelling, synthesis and analysis of biorefinery networks

Maria-Ona Bertran

PhD Thesis
November, 2017

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering
Technical University of Denmark

Preface

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Chemical Engineering at the Technical University of Denmark (DTU). The PhD project was carried out at the Department of Chemical and Biochemical Engineering from September 2014 to November 2017 under the supervision of Professor Rafiqul Gani and Professor John M. Woodley.

This project has been funded by a DTU Alliance Scholarship in the framework of the ProBioRefine network. The project has included two short external stays, carried out at the Technical University of Eindhoven with Associate Professor Edwin Zondervan, and at Texas A&M University with Professor M. Nazmul Karim.

Maria-Ona Bertran

Copenhagen, November 2017

Acknowledgements

This thesis is the result of three years of learnings and challenges that I could not have overcome alone. I would therefore like to thank my colleagues, friends and family for helping me turn this into a rewarding and fun experience.

I would first and foremost like to thank my supervisors Rafiqul Gani and John Woodley for giving me the opportunity to carry out this project, for their encouragement and support, and for challenging my ideas in interesting discussions that continuously drove the development of this work. From my very first month as PhD student, they have trusted me and given me so many opportunities to travel the world in order to keep up with the state of the art, present the outcomes of this project, and be part of the international collaboration network ProBioRefine.

At the beginning of this journey, I had the chance to visit Edwin Zondervan's group at the Technical University of Eindhoven. I am very grateful to Edwin, who introduced me to the basics of optimization and allowed me to get to know his lovely students and learn about their projects.

My second research stay took place in Texas A&M University. I would like to thank Nazmul Karim for his hospitality and for helping me establish collaborations with students in his lab. I am also grateful to his students, Alex and Jon for their contributions to our collaboration work and for showing me around College Station. I would also like to thank Stratos Pistikopoulos and the students at GERB for their hospitality and for giving me the opportunity to present and discuss my work before such a brilliant set of researchers.

I would like to thank my former and current colleagues at CAPEC-PROCESS and the KT Consortium for providing an environment to exchange ideas and have fruitful discussions. Moreover, I would like to thank two visiting academics who have been available for collaboration and dicussions, Mauricio Sales-Cruz and Teresa Lopez-Arenas, and PhD student Omar Anaya-Reza.

The programming related to the development of Super-O was carried out by Lei Zhang, to whom I am very grateful for devoting time to this project and for being my go-to

person for discussions on optimization during and after his time at DTU.

Next, I would like to express my gratitude to the master students I have co-supervised during my PhD project: Pau Cabañeros, Alberto Orsi, Ana-Sofia Sanchez-Arcilla, and Anders Jaksland. It has been an awesome learning experience to work with them and I appreciate their contributions to this project.

I would of course like to thank my awesome friends from Barcelona, with whom I started my Chemical Engineering adventure, and from Denmark, for being my adoptive family and for their continuous encouragement. I am especially grateful to Albert and Manolis for proof reading this thesis and providing very valuable feedback and to Deenesh for always being available for a good discussion, for offering a different perspective on my work and for continuously encouraging me to "go for it". I would like to thank Pau for his support through the struggles and for celebrating each small victory with me.

Last, I am extremely grateful to my parents Magda and Xevi, for being a source of inspiration and not only telling me but showing me that getting where you want is only a matter of determination and hard work, and to my sister Blanca for always being there for me, *gràcies*.

Summary

The quest for satisfying the needs of a growing population along with the urge to address environmental concerns require the development of novel and more sustainable production processes that utilize renewable resources, such as biomass. Biorefineries have gained interest, as integrated facilities for the conversion of biomass into chemicals, fuels and energy, because they have the potential to maximize biomass value while reducing emissions.

The design of biorefinery networks is a complex decision-making problem that involves the selection of feedstocks, processing technologies, products, geographical locations, and operating conditions, among others. Unlike petroleum-based processing networks, biorefineries rely on feedstocks that are non-homogeneous across geographical areas in terms of their availability, type and properties. For this reason, the performance of biorefinery networks depends on their geographical distribution and surrounding markets. Moreover, biorefinery research is ongoing and these processes are not well established, which means that a large number of potential technologies are continuously developed and need to be evaluated. This corresponds to large amounts of data being available yet not consolidated, systematized or ready to use. In addition, the political and social contexts change rapidly, which requires means for fast assessments given each specific context.

It is thus clear that methods and tools to address some of the design challenges are required, for example, the synthesis of reactions to convert available biomass-based feedstocks into desired products, the selection of processing routes and technologies from a large set of alternatives, or the generation of hybrid technologies through process intensification. Systematic process synthesis and design methods have been developed for traditional chemical processes. However, although many concepts are still valuable, these methods are not directly applicable to biorefinery networks and need to be further extended.

The main contribution of this work is the development of a systematic framework for synthesis of biorefinery networks that integrates the necessary models, methods and

tools for the problem formulation and solution. The developed framework is generic, hence useful within a large range of applications, being able to manage the complexity of problems including synthesis of chemical process, bioprocess and biorefineries. It is also flexible, namely easily adaptable to different problem types. Moreover, an ontology-based data structure for data management and the implementation of the framework in a computer-aided tool, Super-O, make the solution procedure faster, more efficient and easily accessible by non-expert users. The framework has been applied to three relevant biorefinery synthesis examples in order to test its application range and highlight its key features.

Resumé

For at imødekomme behovene fra en voksende population og på samme tid adressere miljømæssige problemstillinger, er det nødvendigt at udvikle nye og mere bæredygtige produktionsmetoder, der anvender vedvarende ressourcer såsom biomasse. Her er bioraffinaderier, der omdanner biomasse til kemikalier, brændstoffer og energi, interessante fordi de har potentiale til at maksimere værdien af biomassen og samtidig reducere skadelige emissioner.

Design af bioraffinaderier er en kompleks beslutningstagningsproblemstilling, der blandet andet involverer udvælgelse af udgangsmaterialer, procesteknologier, slutprodukter, geografisk placering og kørselsbetingelser. Til forskel fra petroleumsbaserede raffinaderier, anvender bioraffinaderier udgangsmaterialer som er forskellige afhængig af deres geografiske oprindelse, tilgængelighed, type og egenskaber. Af denne grund afhænger ydeevnen af bioraffinaderier af deres geografiske placering og de omkringliggende markeder. Ydermere befinder bioraffinaderier sig stadig på et forskningsstadie og processerne er endnu ikke veletablerede, hvilket betyder at et stort antal af potentielle teknologier løbende bliver udviklet og må evalueres. Dette giver store mængder af tilgængelig data som endnu ikke er konsolideret, systematiseret eller klar til brug. I tillæg, så kan politiske og sociale kontekster ændre sig hurtigt, hvilket nødvendiggør metoder, der kan bruges til at lave en hurtigt vurdering for en given kontekst.

Således står det klart, at der er behov for metoder og værktøjer til at adressere nogle af designproblemstillingerne, eksempelvis udvælgelsen af specifikke procesveje og teknologier ud fra et stort sæt af mulige løsninger, eller skabelsen af nye hybridteknologier via procesintensivering. Sådanne systematisk processynteser og designmetoder er blevet udviklet til traditionelle kemiske processer. Men, selvom mange af disse indeholder værdi, er de ikke direkte anvendelige for bioraffinaderier, og der er derfor behov for en videreudvikling inden for dette felt.

Det primære bidrag til feltet i denne afhandling består af udviklingen af en systematisk metode til design af bioraffinaderier, der integrerer de modeller, metoder og værktøjer som problemformuleringen og løsningen påkræver. Den udviklede metode er generisk og derfor brugbar indenfor en lang række af applikationer, da den kan håndtere kom-

pleksiteten af problemer der inkluderer syntese af både kemiske processer, bioprocesser og bioraffinaderier. Den er også fleksibel, da den nemt kan tilpasses forskellige typer af problemstillinger. Ydermere, med en ontologisk baseret datastruktur for datahåndtering og implementering af metoden i computerværktøjet, Super-O, bliver løsningsmetoden hurtigere, mere effektiv og lettilgængelig for brugere, der ikke er eksperter. Metoden er blevet anvendt på tre relevante bioraffinaderieksempler for at afprøve spandet for dens anvendelsesområde og fremhæve dets primære funktioner.

Contents

List of Figures

List of Tables

PART I

Introduction

Part I of this thesis sets the stage for the development of an integrated framework for biorefinery synthesis.

The motivation and background of the work are discussed, followed by an identification of research gaps and needs and challenges for the framework development. Next, the context for process synthesis is provided, as the first stage in the 3-stage approach. Finally, the scope and objectives of this thesis are stated, along with a description of the thesis structure.

CHAPTER 1

Introduction

This chapter presents an introduction to the research area, that is, the background and identified needs and challenges, which are addressed in this thesis. The main topics that provide a background for this work lie in the overlap between process synthesis and synthesis-design of biorefineries. Additionally, mathematical programming and development of computer-aided methods and tools are important topics related to this work.

1.1 Process synthesis

Process synthesis is the step in the preliminary stage of process design that consists in creating the process flowsheet by selecting the component parts and their interconnections [1]. An updated definition of process synthesis is given by Babi et al. [2]:

> "*Process synthesis is to find the best processing route, among numerous alternatives for converting raw materials to desired products subject to predefined performance criteria.*"

A variety of requirements (economical, environmental, energy-saving, optimal resource management, etc.) drive industries to develop new reaction paths to utilize raw materials, intermediate by-products and waste in a more efficient manner towards the development of sustainable processes [1, 3]. The sustainable synthesis problem is represented in Figure 1.1.

This is a very complex and relevant task that involves decision making at many different levels. Therefore, the development of systematic methods for solving synthesis problems has been addressed by various authors. Process synthesis reviews have been published by Nishida et al. [1], Barnicki et al. [4], Li et al. [5], Westerberg [6], Cremaschi [7], and Chen et al. [8]. Available methods for process synthesis can be classified as: (i) heuristics or knowledge based, (iii) optimization based, and (iii) hybrid.

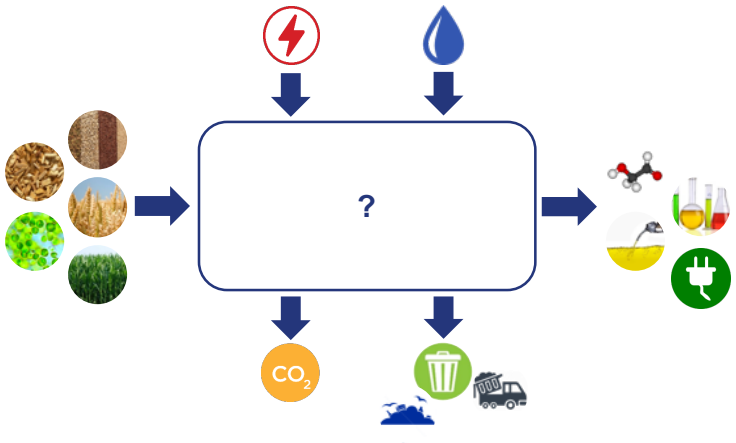


Figure 1.1. The sustainable synthesis problem: generate a flowsheet to convert raw materials to products while using resources (energy, water) efficiently and producing as little waste and emissions as possible.

In heuristics or knowledge based synthesis methods [9, 10], rules derived from prior knowledge and experience are applied. Methods belonging to this class consist in enhancing an existing flowsheet performing one improvement at a time [11, 12, 13]. These methods are limited to exploring solutions that are close to the starting flowsheet. Therefore, hierarchical decomoposition methods have been developed to generate or improve existing flowsheets by dividing the synthesis problem into a series of sub-problems and solving each of them sequentially using heuristics and knowledge [9]. These methods are simple and fast to apply, however, the heuristic rules can be contradictory and the decomposed nature of the solution approach fails to account for interactions between sub-systems (*e.g.*, reactor network and separation network) and decision layers.

Optimization based methods [14, 15] have been developed to overcome the above-mentioned challenge, as these decision making approaches allow the simultaneous explicit consideration of all types of interactions within the system under consideration. These methods consist in representing the solution space a finite set of alternative flowsheets, formulating the decision making problem as a mathematical optimization, and solving the optimization problem. This approach is very powerful due to the ability to evaluate large numbers of alternative solutions, explicitly considering all types of interactions. However, the main drawbacks of this approach are the mathematical complexity of the problem and the fact that only solutions embedded in the superstructure can be obtained.

Hybrid methods combine characteristics from both of the aforementioned approaches,

where a series of sub-problems are formulated and solved [4]. The structure is similar to knowledge based approaches, yet knowledge is replaced by insights related to the behavior of chemicals in the process [16, 17]. The information from insights assists in narrowing down the search space, hence obtaining a smaller sub-problems, which can be solved mathematically.

1.1.1 Optimization-based process synthesis

Available process synthesis methods based on the optimization approach are reviewed in this section. Moreover, the key elements of these approaches are discussed (superstructure representation, models, solution, etc.).

A literature review using keywords within the topic (using the keywords "superstructure optimization" and "mathematical programming & process synthesis") has been performed in Scopus search engine. It should be noted that only peer-reviewed journal papers are included (that is, not considering conference papers, book chapters or other items) and the search was limited to the works within the Chemical Engineering area, as some keywords may be used differently across disciplines.

Superstructure optimization The term superstructure optimization has been present in the Chemical Engineering literature for over three decades. According to a literature search in the Scopus database, it first appeared in the title, abstract or keywords of an article in the field of Chemical Engineering in 1983. The evolution of number of documents on the topic over time is shown in Figure 1.2. A steep increase in the number of published documents on the topic is observed over the years going from around 10 items per year in the 90's to over 60 documents published annually between 2010 and 2015.

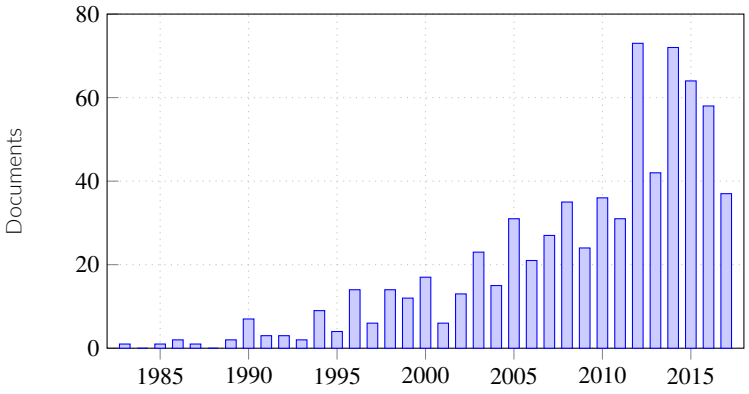


Figure 1.2. Scopus documents by year: results from literature search of articles and review articles containing superstructure optimization in title, abstract or keywords (Search performed late-2016).

Mathematical programming and process synthesis The terms mathematical programming and process synthesis have been used in Chemical Engineering to refer to the optimization-based approach to generate process flowsheets. The number of documents on these topics over the years are displayed in Figure 1.3. It is observed that this terminology has been used for a longer period of time. A steep increase in the number of published works on this topic was observed in the 90's, and has become steady ever since, oscillating between 10 and 15 documents per year. The interest increased again around 2013-2015, which corresponds also to an increase in published works on superstructure optimization. Note that the drop in 2016-2017 is due to the fact that the search was performed before the end of 2016.

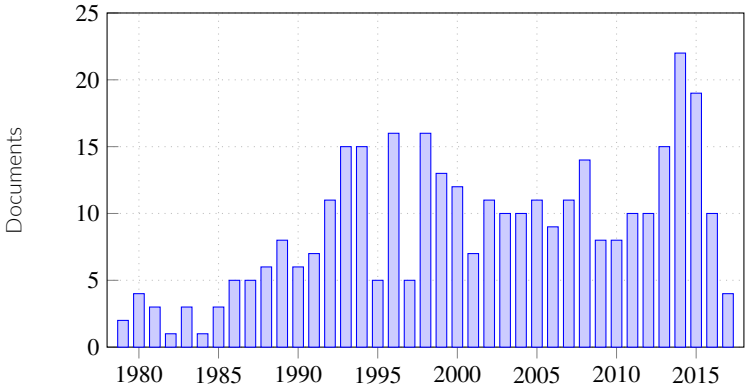


Figure 1.3. Scopus documents by year: results from literature search of articles and review articles containing mathematical programming & process synthesis in title, abstract or keywords (Search performed late-2016).

Generic frameworks for synthesis of processing networks have been proposed [18, 19, 20]. Optimization-based synthesis methods generally consist in three main steps: (i) generating a superstructure of alternatives, (ii) modeling these superstructure, and (iii) solving the optimization problem.

Superstructure representations

The representation of the search space in a process synthesis problem is often done via superstructures. A superstructure is a representation of the topological alternatives in a synthesis problem. Different superstructure representations are known and can be used for different problems. Two well-known superstructure representations are the STN, State-Task Network, and the SEN, State-Equipment Network [18].

In the STN representation, alternatives are represented according to the task that they carry out, whereas in the SEN representation, they are classified in relation to the equip-

ment in which they take place. Due to this variations, alternatives leading to the same set of designs can be represented in very different ways, which yield a different number of alternatives, hence a different problem size, as well as to a different mathematical representation of the problem. Therefore, the use of a generic superstructure representation that can represent all the different alternatives and aids in the modeling step is very important.

Models

The general mathematical formulation of process synthesis is illustrated in Figure 1.4.

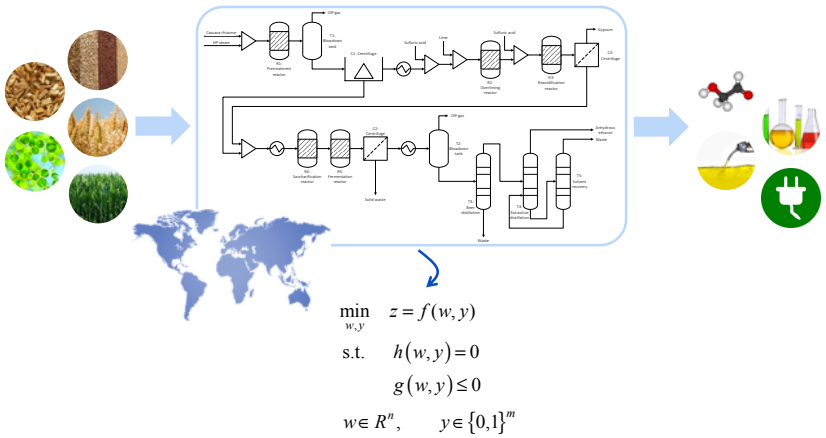


Figure 1.4. Mathematical formulation of the biorefinery synthesis problem.

Specific models can however differ from problem to problem, given that a single superstructure can be represented by different models depending on the desired objectives and the level of detail required [8]. Superstructure models for synthesis span from simple input-output models, in which different phenomena are aggregated, to shortcut models, providing a partially simplified representation of the process units, all the way to rigorous models, where process units are represented in detail including, for instance, mass transfer and equilibrium relationships.

Optimization methods

Optimization algorithms are continuously being developed in order to cope with the complexity of models such as those required for process synthesis. The combinatorial nature of the synthesis problem often yields to a large number of alternatives to be explored in the synthesis stage. Therefore, the problem size is one of the challenges for their solution. Another issue is related to the complexity of the models. Modeling

of unit operations and interconnections in processing networks may lead to complex nonlinear models, for which most algorithms guarantee only global optimal solutions for convex cases. If nonconvexities are present, then rigorous global optimization techniques are required [21, 22]. Grossmann et al. [23] present a comprehensive description of different solution algorithms for the process synthesis problem.

Software tools

The process systems engineering (PSE) community has developed and applied a large number of computer-aided tools to a large number of problems of industrial and academic relevance within the areas of process and product design. Process simulators [24, 25, 26] are widely used both industrially and in academia. Tools for economic, environmental and sustainability analysis are also available, either embedded in the aforementioned packages, or separately [27, 28, 29]. Available tools for process synthesis are mostly limited to applying a decomposition-based approach and/or heuristic rules. For example, ICAS is a integrated software that consists of multiple toolboxes to solve different aspects of the synthesis-design problem [30]. PROSYN is a knowledge-based synthesis tool that guides through synthesis decisions using a decomposition approach and heuristic rules [31, 32, 33]. Very few attempts have been made towards the development of a synthesis tool using an optimization approach, an example is MIPSYN, which was developed by [34] based on a previous version named PROSYN [35, 36].

Overall, despite the above-mentioned attempts, a commercial synthesis tool based on a mathematical optimization approach is currently not available, hence optimization-based synthesis is not widely applied in industrial practice given the lack of expertise. A commercial synthesis tool would be instrumental towards bringing this class of methods into other areas such as industry or government.

1.2 Synthesis and design of biorefineries

The shift from crude oil to biomass feedstocks for the production of chemicals and fuels requires suitable methods and tools for synthesizing, evaluating and selecting process alternatives, and for designing promising processes. Systematic frameworks with a mathematical programming approach have been applied to synthesis-design of biorefineries by various authors.

Optimization-based process synthesis of biorefineries Figure 1.5 shows the evolution of the number of articles (in the Chemical Engineering field) in the topic of superstructure optimization of biorefineries. The relatively low maturity of the topic is observed along with the growing trend in the number of publications from 2011, where only 2 docu-

ments where published, until 2016, with 14 published articles so far. A detailed review of publications in the topic in 2014–2016 is given in Table A.1 on page 184. The articles are listed according to the following characteristics: scope; feedstock, products and technologies; objective function and size; main results; models, problem type, and solver.

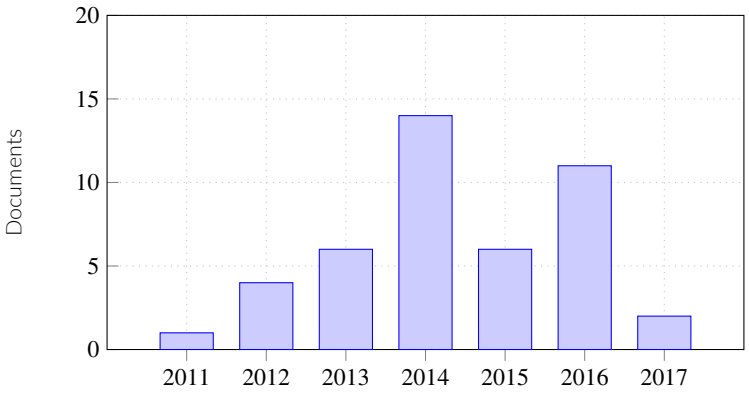


Figure 1.5. Scopus documents by year: results from literature search of articles and review articles containing biorefinery & superstructure in title, abstract or keywords (Search performed late-2016).

The most common feedstocks in works published in 2014-2016 are lignocellulose and algae, popular products include ethanol, biodiesel and succinic acid, and the technologies appearing the most in the reviewed literature are thermochemical and biochemical. Economic objective functions are most often used, which include gross operating margin (GOM), total annualized cost (TAC), net present value (NPV) and gross operating profit (GOI), among others. However, environmental objective functions are also used in the reviewed literature, mostly in the form of global warming potential (GWP). Finally, yield-based objective functions are less common but also used.

Most of the publications related to superstructure optimization of biorefineries solve the optimization problem to determine the optimal network of processing technologies (i.e. flowsheet), set of feedstocks and product portfolio. However, a significant number of them also solve to determine a set of optimal operating conditions. Some of the publications include sensitivity analysis, optimization under uncertainty, and/or multi-objective optimization.

In terms of modeling, most of the reviewed works use shortcut or input-output models, whereas only one of the publications reports the use of surrogate models and another one uses fully rigorous models. Depending on the representation of process models as well as the objective function or constraints, the obtained optimization problem is linear or nonlinear. Out of the 21 publications reviewed, 9 solve the MINLP form of the

problem, 7 solve an MILP, and the rest use an NLP or LP formulation, or do not specify it. Note that if the problem was first formulated as MINLP and then reformulated to obtain an MILP form, it is counted as MILP, since this is the form in which it was solved.

A review of previous works applying and developing methods for optimization-based biorefinery synthesis reveals that most authors have been focused on solving specific problems concerning a set of feedstocks, products or a specific type of biorefinery. For example, the production thermochemical fuels from biomass was studied by Gassner et al. [37], bioethanol production from lignocellulosic biomass was dicussed by Martín et al. [38], a multi-product biorefinery problem including gasoline production was developed by Zondervan et al. [39], and biodiesel production from microalgae was addressed by Rizwan et al. [40]. Recently, González-Delgado et al. [41] suggested a combined method for the synthesis of a microalgae-based biorefinery, starting with identification of promising pathways via a hierarchical approach and continuing with superstructure optimization. The synthesis and design of downstream processes in biorefineries was addressed by Corbetta et al. [42] through an interface between an optimization environment and a process simulator. In terms of generic approaches to modeling, a library of equation-based models for design of biofuel production processes using superstructure optimization was developed by Martín et al. [43].

The synthesis of biorefinery supply chains has also gained interest in the last years. Reviews on biomass supply chain Iakovou et al. [44], Nikolopoulou et al. [45], Sharma et al. [46], Elia et al. [47], and Garcia et al. [48] provide an overview of the main issues and developments in biomass supply chain. These include coordination between various stakeholders, accounting for environmental and sustainability issues, and applying both deterministic and stochastic methods, to account for uncertainties.

Biorefinery databases

Data management has been highlighted as a key aspect of biorefinery synthesis-design, which should cover new and available biorefinery data, including feedstocks, technologies, intermediates, and products as well as economics, market data, locations. Currently available databases and ontologies in the context of biorefineres are listed in Tables 1.1 and 1.2. Available databases are listed in terms of name of each database, category of the content (material and/or technologies, where material refers to biomass, intermediates, and products and technologies refers to transformation processes), details about the content and reference. Note that simulation libraries or other compilations of files have not been considered in this analysis. Only databases that have been especially developed with the objective to organize biorefinery-related data are included.

Available databases and data management approaches for biorefineries are limited to a

specific aspect within biorefineries, such as biomass-based feedstocks, or a specific type of technologies. A database that covers the entire range of data necessary for solving synthesis problems would be instrumental in assisting the solution of these problems.

Although important steps have been made, there is still the need for a generic integrated framework that can address a wide range of problem types within the area and assist in the fast evaluation of alternatives in the early stages of design. Unlike crude oil, each different biomass source provides a different feedstock in terms of chemicals, their composition and properties, and even the same feedstock varies in its characteristics based on the region and climate. Therefore, unlike the optimal petrochemical refinery, the optimal biorefinery network problem needs to have location-dependent solutions. These issues need to be considered during the development of such methods and tools. Moreover, the interaction between processes and the supply chain should be taken into account already from the synthesis stage.

1.3 Needs and challenges

The development of systematic methods and tools to assist the complex tasks involved in the design of novel processing networks, such as biorefineries, has been identified as a key need for which a PSE approach is perfectly suited. The decision making involved in early-stage design of biorefinery networks involves multiple factors, which need to be accounted for simultaneously, in order to capture synergies between them.

A method to determine the optimal configuration biorefinery networks with respect to the pre-defined performance criteria needs to provide answers to the following questions (unless the answer is given by the definition of the given problem):

Q1. Where should each part of the biorefinery network be located?
Q2. Should the network be centralized or distributed?
Q3. Which raw materials should be used?
Q4. What is the optimal product portfolio?
Q5. Which units (process intervals) should be included in the flowsheet to convert raw materials to products?
Q6. How should the selected process intervals be interconnected?
Q7. What are the optimal capacities of the selected process intervals?
Q8. Where should raw materials be sourced from?
Q9. Which markets should be considered for the products?

The answers to questions Q1 and Q2 refer to the geographical configuration of the process network. The answers to questions Q3, Q4, Q5, and Q6 provide information on the structure or topology of the process flowsheet, since they cover the selection of in-

Table 1.1. Scopus biorefinery ontologies, databases, and repositories .

Database	Category	Content	Ref.
QbD Database	Technologies	QbD based – representing 2882 potential biorefinery routes. Reaction yield, utility consumption, and separation efficiency	[49]
Bioethanol Feedstock Geospatial Database (BFGD)	Material	Amount and location of U.S. corn harvested for use as U.S. bioethanol feedstock. To evaluate environmental impacts of bioethanol and identify conservation priorities. Data for years 2005–2010	[50]
Sourcing and conversion database	Material & technologies	Aid biomass sourcing and conversion decision-making. Origin, logistics, technical suitability (in this case for a proprietary organosolv pre-treatment process) and policy and other risk attributes of the system. Feedstock, logistics, policy mechanisms, cost	[51]
Lignin compounds database	Material	Lignin compounds. Softwood (SW), hardwood (HW) and non-wood technical lignins and to compare then lignins derived from key biorefinery technologies	[52]
Waste Biorefinery Ontology	Material & technologies	Chemistry, process models, technology specifications, feedstock characteristics, etc.	[53]
BiOnto	Material & technologies	Biomass types and composition, as well as biorefining technologies, classifying biomass in 5 different ways: i) processing characteristics, ii) chemical composition, iii) physical properties, iv) existing standards, and v) waste and residue based classification.	[54]
Biorefineries web repository	Material & technologies	Biomass conversion pathways to products, models (technology, physical properties, models, cost) and data (experimental or predictive).	[55]

Table 1.2. Biorefinery databases identified through a Google search.

Database	Category	Content	Ref.
BIODAT Database	Material	Biomass, biofuels and ashes for the use of biomass as fuel. Fuel properties, chemical analyses, physical properties, ash properties	[56]
BIOBIB	Material	Biofuels & Thermal utilisation of biofuels analysed by standard analytical methods. Ultimate analysis and combustion data	[57]
BERC Community-Scale Database	Other (facilities)	Community-scale facilities that use modern wood heating or combined heat and power (CHP) systems in US, Canada and Europe. To assist stakeholders in understanding the current demand for wood fuels as an energy source and the viability and success of modern wood heating projects	[58]
Wood2Energy	Other (facilities)	Geospatial database on North American wood to energy conversion facilities. Community through industrial scale systems in US and Canada	[59]
Phyllis2	Material	Composition of biomass and waste: analysis data of individual biomass or waste materials or average values for a group of materials.	[60]
Bioenergy KDF Library	Material & technologies	Publications, data sets, and models specifically related to bioenergy production, distribution, delivery, and end use.	[61]
SAHYOG Biomass Database	Material	An overview of all available biomass potential from different categories of biomass and biowaste resources in EU 27 Member States and in India.	[62]
Bioenergy Feedstock Library	Material	Database for physical, chemical and conversion performance characteristics of biomass feedstock.	[63]

puts/outputs, units and interconnections, while the answers to question Q7 indicates the optimal values of design parameters. The answers to questions Q8 and Q9 cover supply chain aspects. Both discrete and continuous decision making are involved in answering to the above questions, for instance, the selection of process units is a discrete decision whereas the determination of the optimal capacity for a given unit is a continuous decision.

These decisions need to be made subject to the overall process model and design constraints while taking into account supply and demand requirements, transportation through the network, location-dependence of parameters, and specific characteristics of biomass and biorefinery processes.

Systematic approaches to solve such problems are required. Based on the above needs, a biorefinery synthesis method should be able to: (i) generate and represent all possible alternative networks; (iii) model all alternatives; (iv) evaluate and select from the set of alternatives based on pre-defined performance criteria; and (v) manage the relevant data for each problem at hand.

CHAPTER 2

The 3-Stage approach

This chapter provides the context for the work of this thesis, which focuses on the development of a framework for biorefinery synthesis. The synthesis stage is the first of three stages in the 3-stage approach, a concept presented by Babi et al. [2] and further described in Bertran et al. [64].

The mathematical formulation of the design problem is first presented, in order to understand the problem structure and the role of each step towards the solution of the overall problem, which is often too complex and large to be solved directly. That is, the overall process design problem including flowsheet selection, equipment design, sizing, costing, sustainability analysis, process intensification, and others, is extremely complex.

The 3-stage approach is then presented, in which the overall design problem is addressed in three sequential stages in order to manage its complexity. Systematic methods can be implemented in each of the stages. The framework developed in this thesis represents a method for the first stage (synthesis). However, once Stage 1 is completed and a processing route is identified, the selected route is to be further designed and analyzed using more rigorous models and detailed analyses and may be further improved in Stage 3.

2.1 Mathematical formulation of the design problem

Due to its decision-making nature, the synthesis-design problem involves a combination of continuous and discrete variables, hence leading to its formulation as a mixed-integer nonlinear programming (MINLP) problem. In a process design problem, the objective function to be minimized (or maximized) is typically an economic or environmental function given as

$$z = f(w,y)\,, \tag{2.1}$$

where w is a vector of n continuous variables related to processing such as flow rates, operating conditions (temperature, pressure) and equipment design parameters, y is a vector of m integer variables representing discrete decisions, for instance, the existence of a process unit in the optimal process flowsheet, and z is the objective function that represents a performance criterion typically an economic or environmental function. The physical performance of the process is represented by

$$g(w,y) = 0\,, \tag{2.2}$$

where $g(w,y)$ are equality constraints, which may represent mass and energy balances and other process constraints. The design specifications, physical constraints, and logical constraints are given as

$$h(w,y) \leq 0\,, \tag{2.3}$$

where $h(w,y)$ are inequality constraints corresponding to design specifications and other restrictions. Therefore, the overall design problem is expressed mathematically [65, 66] as follows

$$\begin{aligned} \min_{w,y} \quad & z = f(w,y) \\ \text{subject to} \quad & h(w,y) = 0 \\ & g(w,y) \leq 0 \\ & w \in R^n, y \in \{0,1\}^m\,. \end{aligned} \tag{2.4}$$

In Equation (2.4), if the functions f, h, and g are all linear, then the problem is a Mixed-Integer Linear Programming (MILP) problem, otherwise, if any of them is nonlinear, it becomes a Mixed-Integer Nonlinear Programming (MINLP) problem.

2.2 Overview of the 3-Stage approach

In the 3-Stage approach [2], the overall process design problem is decomposed into three sequential sub-problems: (1) Synthesis (superstructure optimization-based process synthesis); (2) Design (detailed design and economic-environmental analyses);

and (3) Innovation (more sustainable design through process intensification and integration). A graphical representation of the 3-stage approach is shown in Figure 2.1, which illustrates inputs and outputs of each stage. The synthesis stage takes raw material, product and technology alternatives and finds the processing route, a feasible design is then determined in the detailed design stage given the processing route. The design becomes the input for the innovation stage, where more sustainable designs are generated. Note that any of the stages can be taken as starting point, provided that the appropriate input data is available. For example, if a flowsheet is available, Stage 2 can be performed directly.

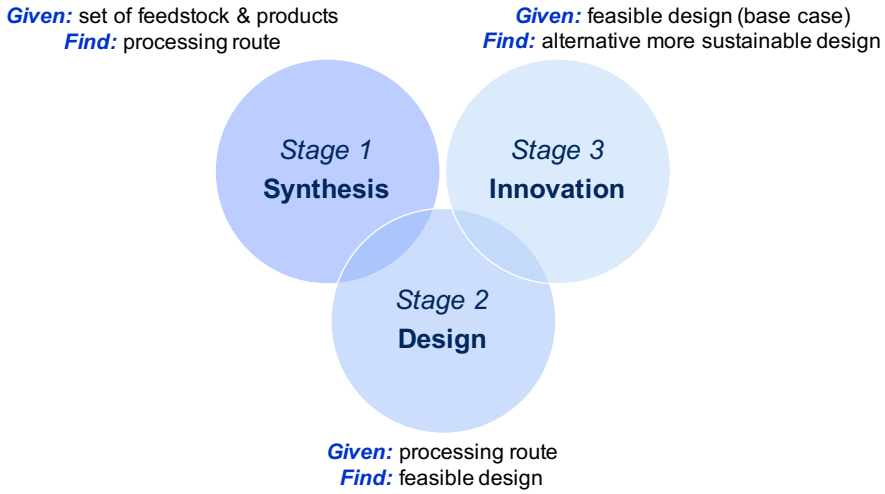


Figure 2.1. The 3-stage approach: (1) synthesis, (2) design, and (3) innovation. Any of the stages can be accessed first for a given problem, provided that the necessary information is available.

Along the three stages, the size of the search space is subsequently reduced, stage by stage, whereas the complexity of the models used is increased from Stage 1 to Stage 3. In this way, qualitatively correct, simpler models are used to select from among a larger number of alternatives in Stage 1, thereby generating a smaller number of feasible alternatives, which are then analyzed in more detail with more rigorous and quantitatively correct models in Stage 2. After that, the detailed design of a feasible processing route is available along with the results of its analyses, so one may decide to stop, or, may use the identified targets for further improvements in the next stage, Stage 3. Here, methods and tools for process synthesis-intensification are applied to determine new, innovative and more sustainable process alternatives. By following this approach, infeasible processing networks are disregarded at Stage 1, therefore avoiding the time-consuming task of performing detailed modeling, simulation and analysis of these alternatives in Stage 2.

The outputs of the 3-stage approach are sustainable process solutions, for which a reference point is needed, which is generated in Stage 1 and designed in detail in Stage 2. Based on this base case, improvements are made in Stage 3 towards a more sustainable solution, for which targets are needed in order to direct the aforementioned improvements. Note that if a processing route already exists, Stage 1 may be bypassed. Similarly, if a processing route with a detailed simulation and/or plant data already exist, Stages 1 and 2 may be bypassed.

The objective function, Equation (2.1) , remains the same across the three stages. However, the problem constraints take different forms. In Stage 1, parameters for each process interval model are given; that is, Equation (2.2) is represented by simpler shortcut models and Equation (2.3) includes superstructure connectivity constraints, logical constraints and supply chain constraints (availability, demand), and only a subset of continuous variables are calculated, corresponding to flow rates and other process variables. In Stage 2, equality constraints, Equation (2.2), are detailed models representing the process flowsheet, including each unit operation, and inequality constraints, Equation (2.3), include design specifications, such as product purity, and equipment constraints. The continuous variables involved in this stage are equipment design parameters, whereas discrete (binary) variables related to the process topology are fixed, hence not calculated. However, discrete variables might appear as design parameters, for example, the number of stages in a distillation column. Finally, in Stage 3, innovative designs are explored with Equations (2.2) and (2.3) and represented by phenomena-based mass-energy balances and constraints. Over the stages, there is a decrease in scale, from the processing interval scale in Stage 1, through the unit operation scale in Stage 2, and until the smaller phenomena scale in Stage 3. A representation of the 3-stage approach is displayed in Figure 2.2.

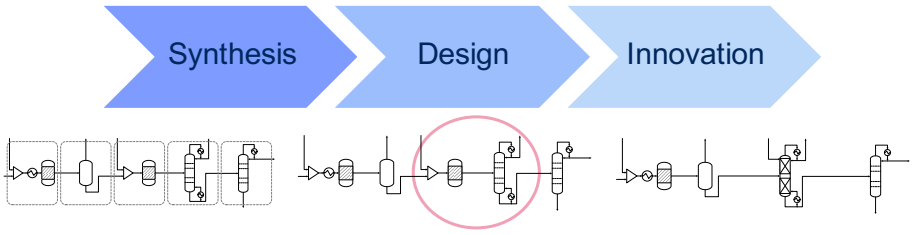


Figure 2.2. Graphical representation of the 3-stage approach: in the first stage, simpler models are used to select a processing network, which is further designed in Stage 2 and analyzed to identify bottlenecks. These become targets for improvement addressed in Stage 3 through, for example, process intensification.

An overview of different problem characteristics, data, models and variables involved in each of the three stages is given by information listed in Table 2.1. Next, details about

each of the stages are discussed in the following sections.

2.3 Stage 1

The objective of the synthesis stage, Stage 1, is to determine the best processing route (process flowsheet) from among numerous alternatives for converting raw materials to products, subject to design constraints and predefined performance criteria. More specifically, the aim is to select the set of raw materials and the set of products, along with their corresponding processing route(s) and the technologies that are able to perform the tasks involved in the processing routes. Various methods have been developed for process synthesis [1], which can be classified in three different types: heuristic-based methods, mathematical optimization methods, and hybrid methods.

In this thesis, a mathematical optimization-based method for process synthesis (Stage 1) is developed and applied to biorefinery processes.

2.4 Stage 2

The design stage, Stage 2, concerns the detailed design and economic-environmental analysis of a given processing route. The outputs of this stage, therefore, include flow rates, utilities, detailed design of technologies and sizing parameters, and results of the analyses performed, that is, economic indicators, sustainability metrics and environmental impacts. These results are used to identify bottlenecks in the process, that become targets for improvement in Stage 3.

As previously mentioned, all stages are mathematically defined by Equation (2.4), with the difference that now more complex models are used as equality constraints and binary variables in terms of process topology are fixed. However, new integer variables might appear concerning discontinuous decisions regarding, for instance, design of equipment (*e.g.*, number of trays in a distillation column, number of compression stages, among others). In terms of methods, traditional process design methods and knowledge are used for design [67]. The main steps followed in Stage 2 are: detailed design and simulation, process optimization, and analysis.

2.5 Stage 3

The innovation stage, Stage 3, is the final stage of this approach. It takes as input the detailed process design and targets for improvement, which have been generated from prior analyses. The outputs are innovative and more sustainable designs for the given

Table 2.1. Characteristics, input-output data, optimization problem form, and problem variables for each of the three sub-problems (stages).

		Stage 1: Synthesis	Stage 2: Design	Stage 3: Innovation
Problem characteristics	**Number of alternatives**	Large	Medium	Small
	Model complexity	Low	High	High
	Scale	Interval	Unit operation	Phenomena
Data flow	**Input data**	Process parameters (such as fixed conversion, operating conditions, separation recovery)	Process topology (flow sheet), mass balance, energy balance	Base case design, process bottlenecks
	Output data	Mass balance, energy balance	Equipment design parameters, economic and environmental metrics	More sustainable design with improved metrics
Optimization problem form	**Objective function**	Economic or environmental	Economic or environmental	Economic or environmental
	Equality constraints	Shortcut models	Detailed process model	Phenomena-based process model
	Inequality constraints	Superstructure connectivity, logical, supply chain	Design specifications, equipment constraints	Phenomena-based intensification constraints
Problem variables	**Continuous**	Flow rates, heat duties	Equipment design parameters (e.g. size, operating conditions)	Intensified equipment design parameters
	Integer	Process topology (existence or non-existence of intervals)	Equipment design parameters (e.g. number of stages)[a]	Intensified equipment design parameters[a]

[a] Note that the topology is fixed after Stage 1.

process. The base line for improvement is the base case, that is, the processing route designed in Stage 2 (or obtained from other sources).

Improvements can be achieved in multiple ways, one of them is through process intensification (PI) and process integration. In order to come up with truly innovative solutions, the method proposed by Lutze et al. [68] and extended by Babi et al. [2] can be applied, which breaks processing units into phenomena and, from a lower aggregation level, generates novel unit operations alternatives, including intensified operations.

CHAPTER 3

Thesis scope

The need for a systematic and integrated approach for the synthesis of biorefinery networks has been identified in Chapter 1 as a gap in the current state of the art. Needs and challenges regarding this issue have also been discussed. Moreover, the 3-stage approach to process design is presented in Chapter 2 as the context of said framework. This chapter states the objectives and structure of this thesis.

3.1 Background & motivation

Finding novel and more sustainable production systems is a key step towards addressing the grand challenges of energy, water, environment and food currently faced by modern society. These challenges are driven by the global population growth, expected to reach 9.6 billion by 2050 [69], which is leading to an increase in the demands of food, water, fuels, and commodity chemicals, among others. Moreover, the environmental impacts caused by consuming fossil fuels need to be addressed promptly through the consumption of more renewable resources. Significantly better and/or new processing routes are needed to, just to name a few, convert available resources to useful products, recycle unused material, and reprocess used material, with the aim of satisfying the current and future needs in an efficient and profitable manner without negatively impacting the society and the environment. From a process systems engineering (PSE) point of view, the wide application range and industrial relevance of the synthesis-design problem provides opportunities to develop and employ systematic and generic solution approaches.

To address some of these challenges, the conversion of biomass to chemicals, fuels and energy in biorefineries is receiving increasing attention, because of their potential to maximize biomass value while reducing emissions.

A processing network is a set of processing units and interconnections that convert raw

materials into desired products. In a process system, as raw materials are consumed through chemical, physical, and biological transformations to yield desired products, additional resources are consumed, such as water and energy, while waste and emissions are generated. The objective of process synthesis is to systematically generate alternative processing routes and select the design whose configuration and parameters minimize (or maximize) a pre-defined performance indicator (typically economic and/or environmental) [70]. This class of problems involves the generation and representation of all possible alternatives; the modeling of all alternatives; a means for evaluation and selection of the optimal design; and the management of large amounts of data from various sources.

Thee design of biorefinery networks is a complex decision-making problem that involves the selection of feedstocks, processing technologies, products, geographical locations, and operating conditions, among others. Unlike petroleum-based processing networks, biorefineries rely on a set of non-homogeneous feedstocks, that are distributed geographically and exhibit different characteristics across locations and seasons. Consequently, the performance of biorefinery networks depends on their geographical distribution and surrounding markets, which needs to be taken into account for the early stages of decision making. Moreover, the area of biorefinery research is not yet mature and developments are ongoing as biorefinery processes are not yet well established, which means that a large number of potential technologies are continuously developed and need to be evaluated. This leads to large amounts of data being available yet not consolidated, systematized or ready to use. Data management is therefore a key challenge in this research area. In addition, the political and social contexts change rapidly, which requires means for fast assessments given each special context. Comprehensive methods and tools to assist the evaluation and decision making have the potential to become instrumental for many players, including industry, academia, and governments.

As part of the overall development of biorefineries, the early-stage synthesis of these facilities is a key step involving strategical decisions such as the selection of appropriate biomass-based raw materials, a network of processing units, a set of desired products, and the geographical configuration of the network. Available methods for process synthesis and design need to be further extended towards establishing processes to convert alternative and renewable raw materials, include recently developed technologies and incorporate new design objectives and constraints (*e.g.*, sustainability).

An integrated business and engineering framework was developed by Quaglia et al. [20] consisting of a step-by-step procedure and integrating different methods and tools including a generic model representing the superstructure based on processing intervals. This framework has been further developed and extended in this PhD project.

3.2 Objectives of the PhD project

The aim of this project is to develop and apply a framework for synthesis of novel processing networks, including biorefineries. The framework needs to account for the interactions between various decision levels, thus it should be optimization-based. It needs to provide an appropriate representation of alternatives, and a generic and flexible approach to modeling the evaluation and selection among said alternatives. Moreover, suitable optimization algorithms should be integrated in the framework. In addition, management of knowledge in this step is a key challenge with needs to be addressed so as to enable storage, retrieval, and reuse of relevant data. These needs are represented in Figure 3.1.

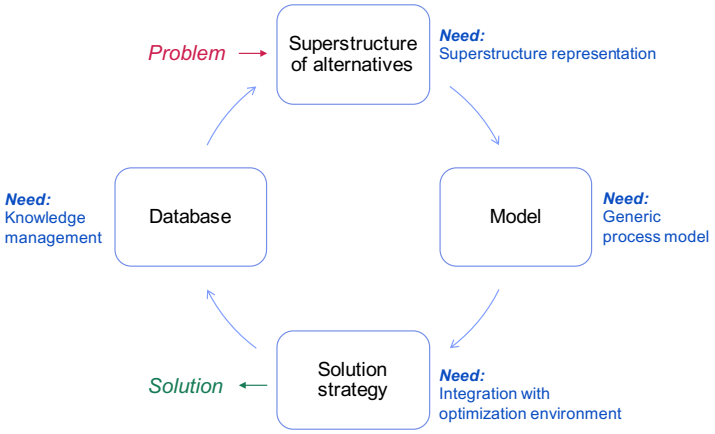


Figure 3.1. Illustration of the key needs for a synthesis framework for novel processing networks.

The framework for processing networks synthesis developed by Quaglia [71] is taken as starting point, to be adapted and extended towards its application to the area of biorefinery synthesis-design. To this end, unique characteristics of biorefinery networks and challenges in their synthesis-design need to be identified in order to target areas for extension of the framework. Moreover, a data structure needs to be developed so as to address data management challenges and provide a platform for systematic storage, retrieval and reuse of biorefinery synthesis data. The framework is to be applied to several examples in the biorefinery area to show some of its possible applications.

3.3 Structure of the PhD thesis

This PhD thesis is divided into four parts: (I) presents an introduction to the research topic, scope, and background, (II) gives a description of the developed framework for

biorefinery network synthesis, (III) consists of application examples to show the application of the framework, and (IV) provides conclusions and future work.

Part I consists of three chapters. Chapter 1 is an introduction to process synthesis, available methods and their application to the design of biorefineries. Chapter 2 presents a three-stage approach to process synthesis, the first stage of which is the focus to this work. Chapter 3 consists the motivation, needs and objectives for the work in this thesis.

Part II comprises four chapters. Chapter 4 is an overview of the framework in terms of needs, objectives, definitions, and the superstructure representation. In Chapter 5, the modeling approach is presented in terms of various modeling blocks and solution algorithms. The workflow and data flow are provided in Chapter 6, followed by the tools used in the framework, which include Super-O, an interface for process synthesis. The data management approach is discussed in Chapter 7, consisting of the development of a data structure, implemented as a database.

Part III consists of three chapters consisting of application examples in increasing order of complexity. The first example is presented in Chapter 8, which corresponds to a process and product network synthesis problem for the production of value-added chemicals from sugarcane molasses. Chapter 9 is a feedstock, process and product synthesis problem in the area of biodiesel production. In Chapter 10, the production of ethanol from various biomass-derived feedstocks is presented, which corresponds to a location-based feedstock and product synthesis problem.

Part IV is composed of Chapter 11, which presents the conclusions of this project and recommendations for future work.

3.4 Dissemination of the PhD project results

The results presented in this thesis and obtained during the PhD project have been presented in 10 international conferences. Moreover, they have been described in the following publications:

Book chapters

I Maria-Ona Bertran, Alberto Orsi, Flavio Manenti, John M. Woodley, Rafiqul Gani, Synthesis of sustainable biofuel production processes: a generic methodology for superstructure optimization and data management. In: Kopanos, G.M., Liu, P., Georgiadis, M.C. (Eds.), *Advances in Energy Systems Engineering*. Springer, Switzerland, **2016**, pp. 651-681.

Journal papers

I Alessandro Rosengart, Maria-Ona Bertran, John M. Woodley, Rafiqul Gani, Attilio Citterio, Flavio Manenti, Process Synthesis for the Production of Sustainable Adipic Acid, Submitted to *Biotechnology for Biofuels*.

II Ramsagar Vororadi, Maria-Ona Bertran, Rebecca Frauzem, Sarath B. Anne, Sustainable chemical processing and energy-carbon dioxide management: review of challenges and opportunities, Submitted to *Chemical Engineering Research and Design*.

III Maria-Ona Bertran, Rebecca Frauzem, Ana-Sofia Sanchez Arcilla, Lei Zhang, Rafiqul Gani, A generic methodology for processing route synthesis and design based on superstructure optimization, *Computers & Chemical Engineering*, in press **2017**, DOI: 10.1016/j.compchemeng.2017.01.030.

Peer-reviewed conference papers

I Maria-Ona Bertran, John M. Woodley, Rafiqul Gani, Location-dependent optimal biorefinery synthesis, *Computer Aided Chemical Engineering*, in press **2017**.

II Alessandro Rosengart, Maria-Ona Bertran, Flavio Manenti, Attio Citterio, John M. Woodley, Rafiqul Gani, Computer aided synthesis of innovative processes: renewable adipic acid production, *Computer Aided Chemical Engineering*, in press **2017**.

III Alexander Sabol, Maria-Ona Bertran, Jonathan P. Raftery, John M. Woodley, Rafiqul Gani, M. Nazmul Karim, Process synthesis framework for the production of high-value intracellular compounds, *Computer Aided Chemical Engineering*, in press **2017**.

IV Maria-Ona Bertran, Rebecca Frauzem, Lei Zhang, Rafiqul Gani, A generic methodology for superstructure optimization of different processing networks. *Computer Aided Chemical Engineering*, **2016**, 38, 685-690.

PART II

Biorefinery synthesis framework

Part II of this thesis presents the biorefinery synthesis framework in terms of its objective and scope, modeling approach, data management and steps.

First, the objectives, key concepts and overview of the framework are presented. Next, a set of generic model blocks are described, which are combined to generate the two main models used in this thesis: (i) a synthesis model, and (ii) a synthesis-location model. The first model is targeted towards flow sheet synthesis problems, while the second one is used to solve location-dependent synthesis problems. These models require large amounts of data, especially when formulating and solving larger synthesis problems. An efficient system for data storage and reuse is therefore necessary. The data management for biorefinery synthesis problems is then presented in terms of the developed data structure, its implementation in the Biorefinery database and methods for data storage and retrieval. Last, the step-by-step workflow and data flow are described in detail. The tools used in the framework are presented and its integration in a software implementation of the framework is reported.

CHAPTER 4

Framework overview

A three-stage approach to process design has been presented in Chapter 2. The approach consists in decomposing the design procedure into three sequential stages to manage the complexity: (1) synthesis (processing route synthesis), (2) design (detailed design and economic-environmental analysis), (3) innovation (more sustainable design through process intensification and integration). A framework for stage 1 (synthesis) especially targeted at biorefinery networks is described in this chapter. First, the objectives of the biorefinery synthesis framework are presented, along with an overview. Next, key concepts are defined and, finally, a description of the superstructure-based representation of alternatives is given.

4.1 Objectives of the framework

The synthesis problem definition is to determine the best processing route (process flowsheet), including feedstock mix and product portfolio, from among numerous alternatives for converting raw materials to products, subject to design constraints and predefined performance criteria. The general objective of a framework towards the systematic formulation and solution of biorefinery synthesis problems is to support strategic (long-term: *what to produce?, what to produce it from?, where to produce?*) and tactical (medium-term: *how to produce it?*) decision making involved in the early-stage design of biorefinery processing networks. Therefore, the biorefinery synthesis framework needs to:

Objective 1. Be generic and flexible, allowing its application to a wide range of problems and adapting easily to the specific characteristics of each of them

Objective 2. Integrate the necessary methods and tools for the systematic formulation and solution of synthesis problems

Objective 3. Manage data systematically allowing its efficient collection, storage and reuse

Objective 4. Be easy to use and accessible to non-expert users

4.2 Problem definition

The synthesis problem is defined as follows:

Given an ensemble of feasible flow sheets in terms of:

1. a set of available feedstocks/raw materials
2. a set of products and by-products of interest
3. a set of processing steps that allow the transformation of feedstocks to products
4. a set of technological alternatives, in terms of processing intervals, for each of the steps
5. a set of relevant geographical locations
6. rules regarding the connections between processing intervals
7. supply-demand data for feedstocks and products
8. mass and energy balance data for each technological alternative
9. economic data

Determine the optimal processing network including:

1. the most appropriate processing alternatives (technologies)
2. flow rates across the network as well as internal flow rates in intervals
3. consumption of utilities and chemicals in each processing interval
4. capacity of each technology
5. cost analysis based on the terms included in the objective function
6. values of post-optimality performance indicators (process performance, sustainability, economic)

Ensuring that the following conditions are satisfied:

1. the flowsheet must be capable of converting input (feed streams) to output (product streams)
2. the selected network must provide an optimal value (minimize/maximize) of a pre-defined performance indicator (objective function) from among all the network alternatives in the considered set

4.3 Overview of the framework

The synthesis of biorefinery processing networks requires the selection of feedstocks, products and by-products, processing routes, technologies, and geographical locations. A computer-aided framework is presented in this chapter towards the systematic formulation and solution of this class of problems. The framework is based on a mathematical optimization approach to process synthesis, which generally consists in proposing a representation of a number of alternative process structures, formulating this as an optimization problem and solving the problem to find the optimal structure.

The three key elements in the formulation and solution of synthesis problems using mathematical programming are: (1) a superstructure-based representation of all considered alternatives; (2) a mathematical model of the superstructure; and (3) a solution strategy for the optimization problem. Moreover, a fourth element is integrated in this work: (4) a database of biorefinery synthesis data, which is built upon a generic data structure and allows efficient collection, storage and reuse of relevant data. A representation of the key elements of the framework is displayed in Figure 4.1, where their connection from problem to solution is shown.

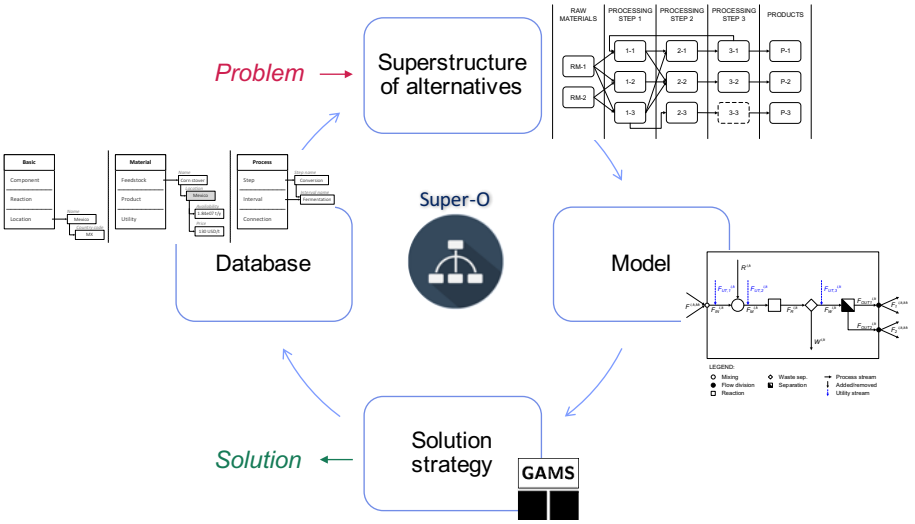


Figure 4.1. The key elements of the biorefinery synthesis framework: superstructure representation, mathematical model, solution strategy and database. The three first elements connect problem to solution, while the fourth element (database) allows data storage and reuse for future problem formulations.

The superstructure representation is described in this chapter (Section 4.5), and all mathematical models and solution strategies are presented in Chapter 5. The framework workflow is given in Chapter 6 along with a description of the software imple-

mentation Super-O. Next, the data structure and database are described in Chapter 7

4.4 Definitions

In this thesis, the concept of *superstructure* is used to represent an ensemble of all possible alternatives being considered in a decision-making problem. A superstructure representation is proposed, where the necessary operations in the process are described as a sequence of *processing steps* and alternatives for each processing step are *intervals*, which can be *processing intervals*, when they represent transformations, or *raw material/product intervals*. Hence *feedstocks* and *products* are represented as intervals belonging to the first and last step, respectively. Moreover, the terms *feedstock* and *raw material* are used interchangeably. In contrast, the term *source* is used for referring to the initial material from which the feedstock is derived. For example, corn stover is a feedstock/raw material and the corn plant is its source.

Intervals are connected to each other via *superstructure connections*, so that a representation of all possible topological solutions is achieved. Superstructure connections are *primary connections* when they connect a primary outlet of an interval with another interval, and *secondary connections* when they connect a secondary outlet with another interval. A primary outlet is the default interval outlet if no separation takes place. When separation occurs, a primary and secondary outlet are present, which are generally defined so that the primary outlet contains the desired product and the secondary outlet contains the byproduct.

The representation described above is known as *Processing Step-Interval Network (PSIN)* and it is suitable for flow sheet synthesis problems in a wide range of applications. Due to the importance of accounting for the geographical location dependence of the solution of biorefinery synthesis-design problems, an extended representation has been developed, named *extended Processing Step-Interval Network (ePSIN)*, which covers the flow sheet synthesis problem as well as supply chain considerations (location selection and transportation). The ePSIN representation requires, not only the definition of steps and intervals, but also of *processing sections*, which are defined as sets of intervals belonging to the same processing plant, processing facility or *Verbund*, hence to be placed in the same location.

A sequence of *processing tasks* may occur within each processing interval. A processing task is defined as an elementary phenomenon incurring physical, chemical or mechanical transformations on the stream(s) being processed. A set of elemental processing tasks are considered in this work, based on those defined by Zondervan et al. [39] and adopted by Quaglia et al. [20]. The *mixing task* refers to the addition of chemicals (*e.g.*, co-reactants, solvents) that are mixed with the process stream. The *reaction task* con-

sists of chemical transformations. The *waste removal task* represents the separation of waste components from process streams, forming waste streams that exit the processing network (*i.e.*, are not further connected to other processing or product intervals). The *separation task* models separations with two stream outlets (primary and secondary) that are both connected to other intervals in the network (either processing intervals or product intervals). This task can be used in various ways, including product/byproduct separations, and the separation of a stream to be recycled. The *utility consumption* task mimics the consumption of heating, cooling, power and other utilites by the interval. Utilities added via this task are not mixed with the process stream.

More definitions of key concepts are given in the Glossary on page 213.

4.5 Superstructure representation

The superstructure representation used in this work is named Processing Step-Interval Network (PSIN), which was initially presented by Quaglia et al. [20]. The basic PSIN representation and an extended version (ePSIN) are presented in this section.

4.5.1 Processing Step-Interval Network (PSIN) representation

A superstructure is a representation of the alternatives comprised within the search space that is considered in a decision-making problem. In this project, a stage-wise superstructure representation is used. The Processing Step-Interval Network (PSIN) representation, illustrated in Figure 4.2, provides an organized platform to represent processing network synthesis problems at multiple scales: from a part of a process to a network of processes. This representation consists of a series of processing steps, represented as columns, and alternatives within each step named processing intervals, represented as boxes (nodes). Alternatives are connected via connections, represented by arrows (directed arcs), meaning that connections are directed mimicking the direction of the flow in the processing network. In the first column (step), raw material alternatives are collected, whereas the last column (step) contains the product alternatives. The remaining columns correspond to the processing steps required for the transformation of raw materials into products.

The presented PSIN representation of Figure 4.2 can be used in the area of biorefinery synthesis to obtain superstructures that represent all the considered alternatives for a given problem with respect to the topology of the network.

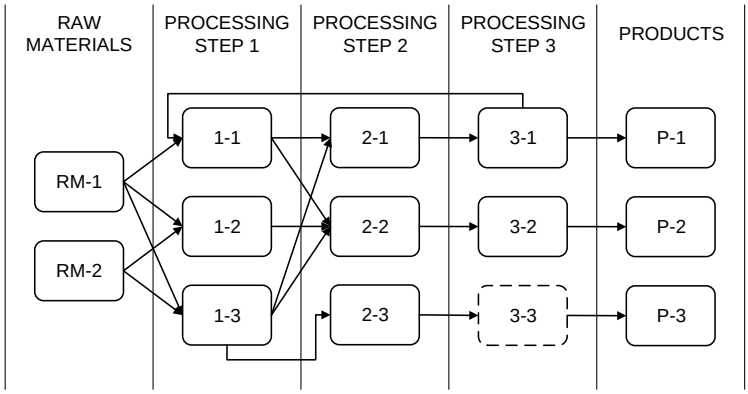


Figure 4.2. Processing Step-Interval Network (PSIN) representation: columns represent processing steps, boxes represent intervals, bypass intervals are highlighted with dashed outline, and arrows represent superstructure connections. The superstructure in the figure contains the following: 3 processing steps, 2 raw material alternatives, 9 processing intervals, including 1 bypass (interval 3-3), and 3 products. In terms of connections, the superstructure has a total of 18 connections, 16 of which are primary connections whereas 2 are secondary, 1 out of the 2 secondary connections is a recycle stream. Secondary connections are identified by their starting point not being placed on the right-hand side of the boxes representing processing intervals, but being placed elsewhere, normally on the top or bottom of the box. When representing in a superstructure with the alternative of bypassing a processing step, two options are allowed: either using a bypass interval, that is placing "empty" interval, or directly connecting intervals from the adjacent steps. Note that processing sections, locations and transportation are not included in this representation.

4.5.2 Extended Process Step-Interval Network (ePSIN) representation

In the area of biorefinery synthesis-design, location plays an especially important role in the decision making due to the fact that the feedstock is not homogeneous across geographical areas. Different types of biomass-based feedstocks are available in different areas and even the same type of feedstock can present variations in its characteristics across locations due to varying meteorological conditions, and soil properties. Therefore, an extended PSIN representation, displayed in Figure 4.3, has been developed so as to represent the selection of geographical locations for feedstocks, processing intervals and products in the superstructure. The selection of locations for feedstocks corresponds to the decision of optimal geographical areas for harvesting, purchasing or collecting of biomass-based feedstocks for their processing in biorefineries. Similarly, product locations correspond to markets for sale of products for their further processing, industrial use or retail. Locations for processing intervals represent options for processing plant locations.

As described in Section 4.4, a processing interval may represent anything from a part

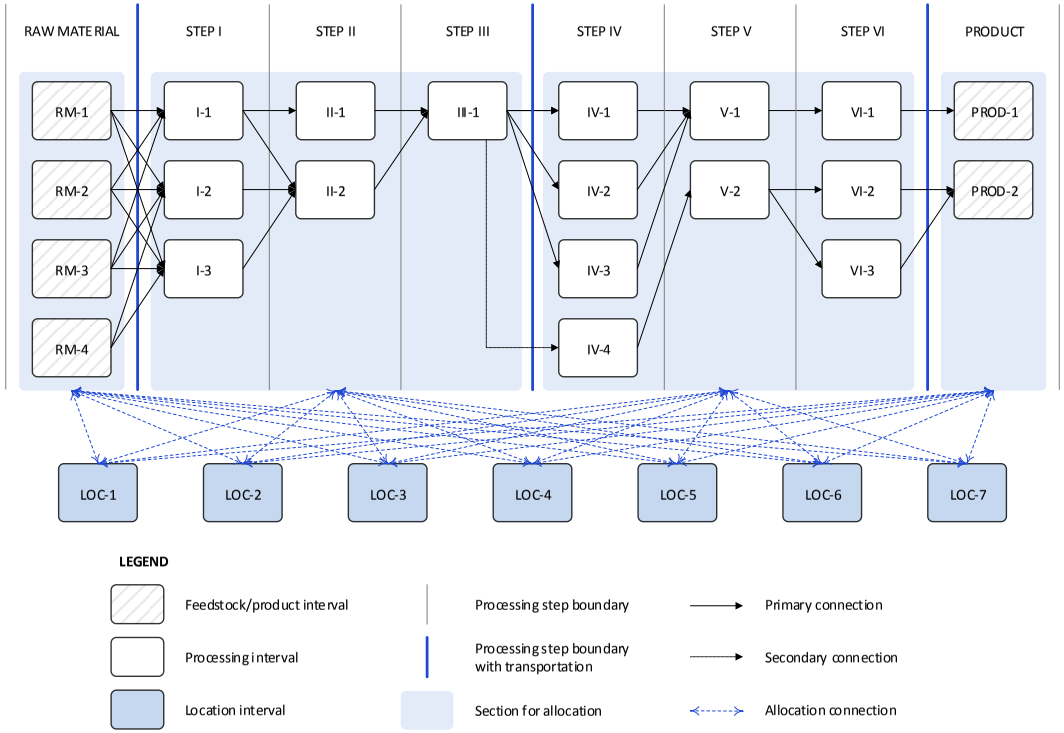


Figure 4.3. Extended Processing Step-Interval Network (ePSIN) representation: columns represent processing steps, boxes represent intervals, bypass intervals are highlighted with dashed outline, and arrows represent superstructure connections. Primary connections are represented as originating from the right-hand side of intervals, secondary connections are depicted originating from the top or bottom of intervals. Large shaded boxes without outline represent sections for location selection, shaded intervals at the bottom represent location alternatives, dashed connections represent location connections between locations and sections.

of, a single or multiple unit operations all the way to an entire process. This allows the representation and solution of problems within a large range of scales, which can give rise to different location selection scenarios. For cases where a single process is represented in the superstructure with alternative processing intervals at the unit operation level, a single location is generally considered for the process. Cases considering entire value chains from raw biomass to end product may involve more than one location for the sub-processes being considered (for example, pretreatment, conversion to stable intermediate, and end-product formulation). Finally, multi-process cases, where an interval represents a whole process, may require the selection of a single location for each of them.

The extended PSIN representation accounts for all the aforementioned scenarios by using the concept of processing sections, which are sets of processing intervals that are to be placed in the same location. Processing sections are user-defined (with respect to their number and content in terms of intervals), however, the first and last sections are automatically defined as the first and last steps, containing the feedstock and product alternatives. The extended PSIN allows the representation of simultaneous process synthesis, location selection and basic supply chain early-stage decision making.

4.6 Problem types

Various problem types can be formulated based on the PSIN/ePSIN representation. Two main problem types are considered in this work: (i) process synthesis problems (represented with PSIN), which are referred to as *synthesis problems*; (ii) process synthesis problems with supply chain considerations (represented with ePSIN), which are designated as *location-based synthesis problems*. It should be noted that storage is not considered in supply chain synthesis in this work and neither is the time variable, that is, planning decisions are not made.

4.6.1 Synthesis problems

Synthesis problem types are shown in Figure 4.4. This class of problems does not include location as ouput, even though inputs may be location-dependent. If this is the case, then problems are solved on a location-by-location basis or with a pre-defined location. Problems derived from the generic synthesis problem include the selection of processing route given a raw material and a product (fixing raw material and product), the synthesis of a process and product network (fixing the raw material), the synthesis of a raw material and product network (given the desired product), the synthesis of raw material, process and product networks (when neither the raw material, nor the process or product are fixed), and the synthesis of feedstock, process and product net-

works via an intermediate product (which, using the nomenclature from Figure 4.4, can be derived from a problem type (c) followed by problem type (b)).

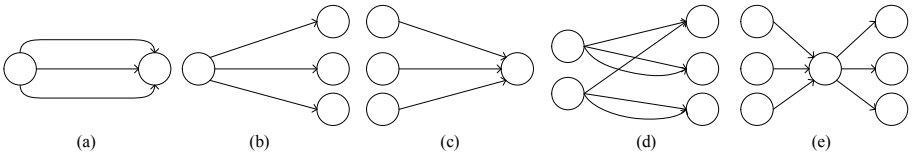


Figure 4.4. Synthesis problem types: (a) route selection problem (fixed raw material, fixed product); (b) product selection problem (fixed raw material); (c) feedstock selection problem (fixed product); (d) feedstock-route-product selection problem; (e) selection of feedstock-product through intermediate.

All the above problems might include by-passes (defined as the possibility of not selecting any interval in a given process step and continuing towards the next one), as well as recycle streams.

4.6.2 Location-based synthesis problems

Location-based synthesis problems are those where selected locations are part of the output of the problem and, therefore, transportation and other supply chain aspects may be included. This class of problems are very relevant to biorefinery networks and account for issues such as the decision of centralized vs distributed networks. Location-based problem types derived from the presented representation of alternatives are illustrated in Figure 4.5.

The concept of processing section gives rise to these different problems. For example, a problem which can be constraint to a single location (see Figure 4.5a), this corresponds to the basic synthesis problem with additonal location selection, that is, the optimal location for the whole network is selected along with raw materials, products and the process topology and configuration. The formulation in Figure 4.5b represents a problem where location selection is performed for raw materials, process and products and various locations are allowed for sources and sinks (*i.e.*, raw material sources and product markets). Next, processing sections can be defined so that the processing network itself is distributed geographically, and optimal locations for each section are determined (see Figure 4.5c). The more distributed case is where each interval represents a section, that is, each can be placed in a different location (see Figure 4.5d). This problem is relevant for enterprise-wide cases and may require that multiple intervals can be selected in each step.

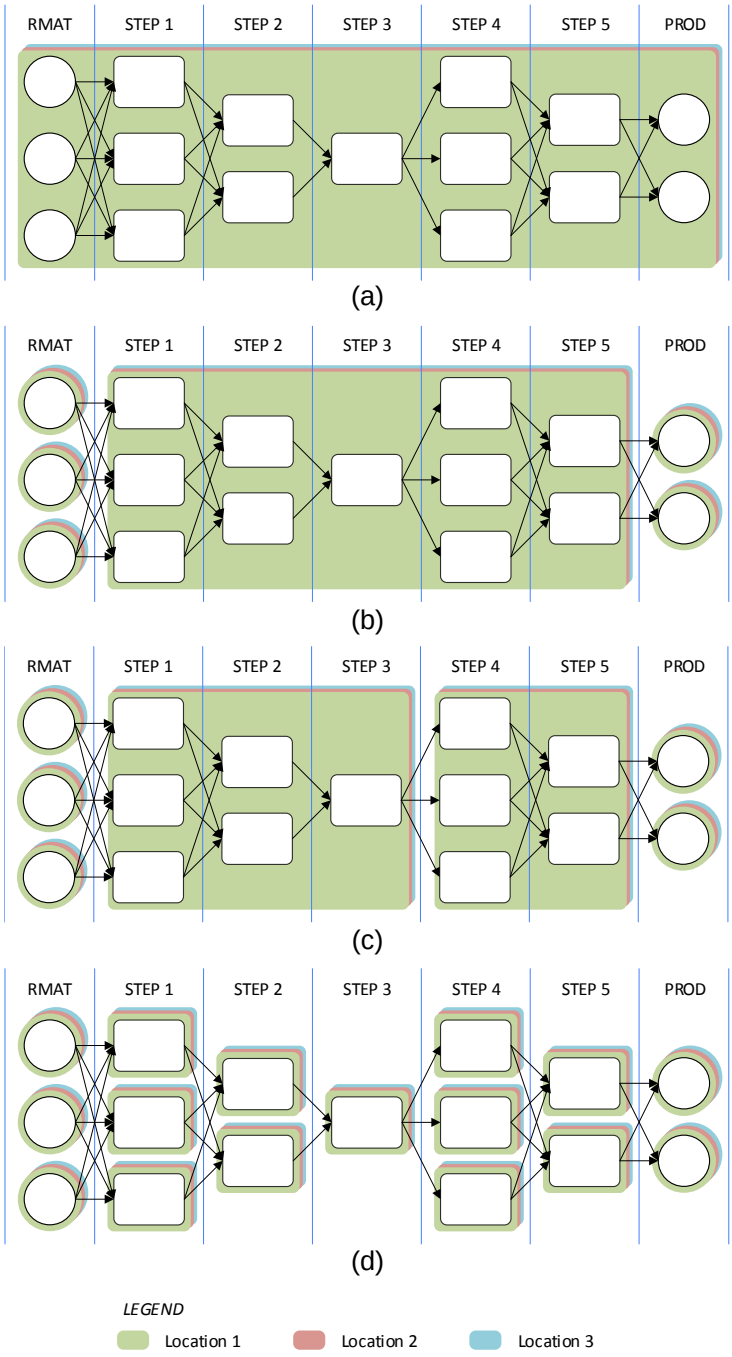


Figure 4.5. Location-dependent synthesis problem types: (a) location-dependent synthesis with location selection constraint to a single location; (b) location-dependent synthesis with location selection for one processing section; (c) location-dependent synthesis with location selection for two processing sections; and (d) location-dependent synthesis with location selection with multiple processing sections. Note that transportation can be accounted for in cases (b) to (d).

4.7 Conclusion

An overview of the developed framework for biorefinery synthesis has been given in this chapter. The key elements of the framework are: a superstructure representation, which is also presented in this chapter, a modeling approach and solution strategies, described in Chapter 5, and a data structure for data management, presented in Chapter 7.

Two superstructure representations are presented: the PSIN and the ePSIN, suitable for synthesis and location-based synthesis problems, respectively. These superstructures (representing both STN and SEN superstructures) can model different problem types. The two main problem types considered in this thesis are synthesis and location-based synthesis. However, variations of these can be obtained, for example a synthesis problem may consist in the synthesis of the process network (where feedstocks and products are known), or can include feedstock, process and product networks. Similarly, location-based synthesis problems might be limited to a single location, or allow distributed configurations. Moreover, blocks in the superstructure may represent units at different scales, ranging from unit operations all the way to entire processing facilities.

CHAPTER 5

Models

Models are a core element of the biorefinery synthesis framework, given its solution approach relying on mathematical optimization. Specific models are normally developed on a problem-by-problem basis for the modeling of superstructures in various application fields, including synthesis of biorefineries. These models often lack flexibility, which complicates their reuse when formulating and solving new problems. Alternatively, generic models are developed to solve a range of problems with reduced time and effort for modeling, especially when used by non-expert users. A generic model for synthesis of processing networks was presented by Quaglia et al. [20], which integrates the generic processing interval representation of Zondervan et al. [39]. This model has been adopted in the biorefinery synthesis framework and extended to suit a wider range of biorefinery-related problems.

This chapter presents the generic model equations in terms of its model blocks: (i) process interval model, (ii) superstructure model, (iii) location model, (iv) transportation model, (v) economic models (capital cost, operating cost), and (vi) objective function. From the presented equations, the two main models used in the framework are presented in two parts: first, the basic synthesis model (for process synthesis problems); and second, the extended synthesis-location model (for location-based synthesis problems). An analysis of these models is used to determine the required model inputs, in other words, parameters that need to be specified for each model block are listed.

Finally, optimization solution algorithms are discussed, which depend on the model form. These algorithms are accessed via its implementation in GAMS [72].

5.1 Overview of models

A generic optimization model for the synthesis of processing networks was proposed by Quaglia et al. [20]. Biorefinery networks have unique characteristics to be taken into account in the synthesis stage, for example, the location-dependency of biomass attributes. Therefore, the available generic model by Quaglia et al. [20] has been adopted and extended for its application to the synthesis of novel processing networks involving the consumption of renewable feedstocks (more specifically, biorefinery networks) The main extensions are:

- Specifying feedstock composition and total flow rate of raw material or product (or upper/lower bounds) is now possible, instead of a specific component-based feedstock flow rate, hence enabling optimization of the capacity.
- Feedstock moisture is added as parameter, which is taken into account when calculating costs.
- Utilities and chemical compounds are placed in different sets, and three utility consumption points are defined, instead of one.
- A location model block is added, allowing the input of location-dependent data and enabling feedstock-plant-product location selection.
- The concept of processing sections is implemented, as sets of intervals that are in a single geographical location.
- A transportation model block is added, that models transportation between sections when these are not placed in the same location.

Overall, the extended model consists of a set of models blocks that allow various formulations relevant for the synthesis of biorefineries. The generic model with its component blocks is shown in Figure 5.1. Equations for each model block are presented in Sections 5.2.1 to 5.2.6.

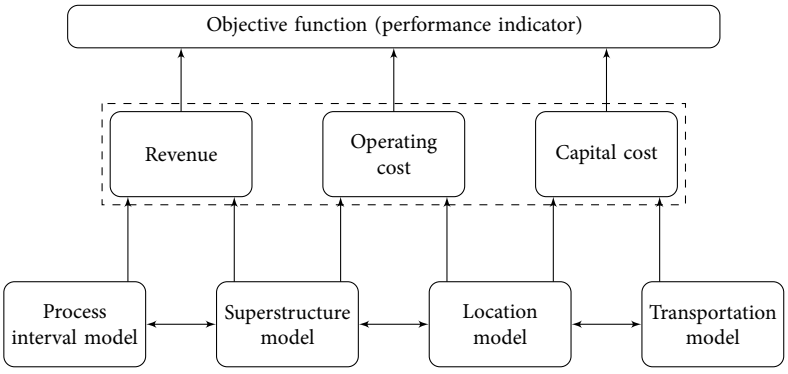


Figure 5.1. Overview of models in the biorefinery synthesis framework.

5.2 Model blocks

5.2.1 Block I: Processing interval model

Processing alternatives are represented as processing intervals, which are modeled using a generic model based on a sequential combination of processing tasks. A set of generic equations is used for each interval representing the sequence of processing tasks, namely mixing, reaction, waste removal and product separation, as well as utility consumption. Multiple inlets to and outlets from the interval are allowed, including recycle streams from downstream intervals and bypasses. A schematic representation of the generic processing interval model is shown in Figure 5.2.

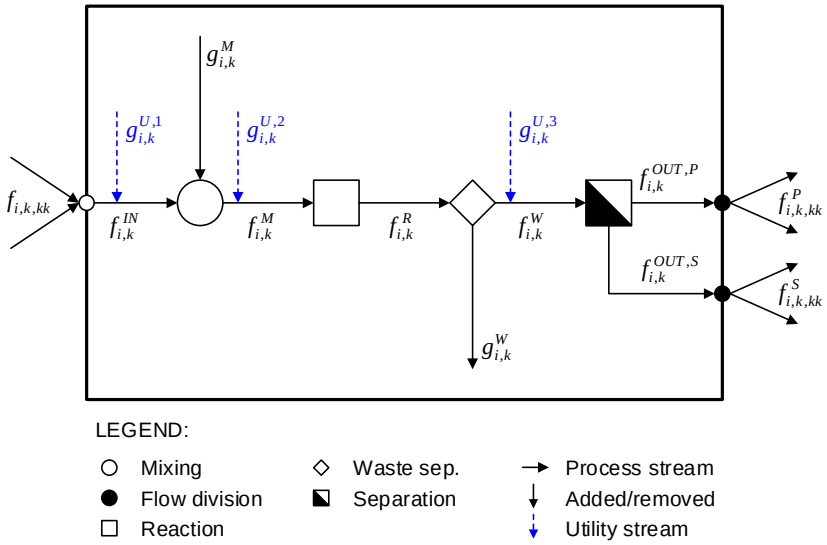


Figure 5.2. Generic processing interval model.

Note that lowercase is used for component flow rates whereas upper case is used for total flow rates. In terms of flow rates, f and F denote flow rates of main process streams going from raw materials to products and/or by-products (represented horizontally in the process interval model scheme in Figure 5.2), g refers to flow rates of incoming or outgoing streams of added chemicals, utilities and wastes (represented vertically in the process interval model scheme in Figure 5.2).

Mixing task

The mixing task models the addition of chemicals (co-reactants, solvents, etc.) to the interval, which are mixed with the main process stream. The flow rate of chemical i added to interval k, $g_M^{i,k}$, is calculated as

$$g_M^{i,k} = \sum_{ii} \left(\mu^{i,ii,k} f_{IN}^{ii,k} \right) \quad \forall i,k \tag{5.1}$$

where the parameter $\mu^{i,ii,k}$ corresponds to the ratio of added compound i in interval k with respect to a reference compound ii and $f_{IN}^{ii,k}$ is the flow rate of reference component ii in the inlet stream of interval k.

The flow rate of component i in interval k after the mixing task, $f_M^{i,k}$, is calculated as

$$f_M^{i,k} = f_{IN}^{i,k} + g_M^{i,k} \quad \forall i,k \tag{5.2}$$

where $g_M^{i,k}$ is the flow rate of chemical i added interval k, as calculated from Equation (5.1).

Reaction task

The flow rate of component i in interval k after the reaction task, $f_R^{i,k}$, is calculated as

$$f_R^{i,k} = f_M^{i,k} + \sum_{r,reac} \left(\theta^{reac,k,r} \gamma^{j,r} \frac{MW^i}{MW^{reac}} f_M^{reac,k} \right) \quad \forall i,k \tag{5.3}$$

where $\theta^{reac,k,r}$ is the conversion of reaction r in interval k in terms of key reactant $reac$, $\gamma^{i,r}$ is the stoichiometric coefficient of component i in reaction r, MW^i and MW^{reac} are the molar weights of component i and key reactant $reac$, respectively, and $f_M^{reac,k}$ is the flow rate of the key component $reac$ in the inlet of the reaction task.

Waste separation task

The waste separation task models separations where one of the streams is further processed, whereas the other one is a waste stream. The waste outlet from this task cannot be connected to other processing intervals in the network.

The flow rate of component i that continues to the next task in the process interval, $f_W^{i,k}$, is calculated as

$$f_W^{i,k} = f_R^{i,k} \left(1 - \delta^{i,k} \right) \quad \forall i,k \notin RAW\,(k) \tag{5.4}$$

where $\delta^{i,k}$ is the waste fraction defined as the ratio of the flow rate compound i in the waste stream to the flow rate of the same component in the inlet stream of the waste

separation task. The flow rate of component i in the waste stream of interval k, $g_W^{i,k}$, is then calculated as

$$g_W^{i,k} = f_R^{i,k} - f_W^{i,k} \quad \forall i,k \notin RAW\,(k) \tag{5.5}$$

where $f_W^{i,k}$ is flow rate of component i in the outlet of the waste separation task of interval k, as calculated in Equation (5.5).

Product separation task

The product separation task is used for modeling of separations with two flow outlets where both outlets are further processed in the network (*i.e.*, they are both connected to other process intervals): a primary and a secondary outlet. Even though these outlets are represented as top and bottom outlets in the graphical representation of Figure 5.2, they do not necessarily need to correspond physically to top and bottom products in the separation being modeled. Generally, the primary outlet is used for the main process stream containing the desired product, while the secondary outlet is used for other streams containing by-products or unreacted raw materials to recycle.

The primary product flow rate of the separation task, $f_{OUT,1}^{i,k}$, is calculated as

$$f_{OUT,1}^{i,k} = \sigma^{i,k} f_W^{i,k} \quad \forall i,k \tag{5.6}$$

where the user-defined parameter $\sigma^{i,k}$ corresponds to the ratio of flow rate of component i leaving the task via the primary product stream to the flow rate of component i in the inlet stream, therefore it is equivalent to the recovery of component i in the primary product stream.

The secondary product stream flow rate, $f_{OUT,2}^{i,k}$, is then by simple mass balance

$$f_{OUT,2}^{i,k} = f_W^{i,k} - f_{OUT,1}^{i,k} \quad \forall i,k \tag{5.7}$$

where $f_W^{i,k}$ is the component flow rate in the inlet of the separation task of interval k and $f_{OUT,1}^{i,k}$ is the component flow rate of the primary outlet of the separation task of interval k, as calculated from Equation (5.7).

Utility consumption task

This task models the utility consumption in each processing interval. The amount of utility consumed in each task of an interval is defined as a function of the total flow rate of process stream in the interval. This is done because along the path within an interval, the flow rate changes due to variations in composition and addition or separation of streams, and therefore different points have been selected in which addition of utilities can be specified. Three different utility consumption points (1, 2, and 3) have been

defined. The total consumption of utility *ut* in interval *k*, $g_{UTIL}^{ut,k}$, is calculated as the sum of the consumption of utilities in each utility consumption point

$$g_{UTIL}^{ut,k} = \sum_i \left(\lambda_1^{ut,k} f_{IN}^{i,k}\right) + \sum_i \left(\lambda_2^{ut,k} f_M^{i,k}\right) + \sum_i \left(\lambda_3^{ut,k} f_W^{i,k}\right) \quad \forall ut, k \tag{5.8}$$

where $\lambda_1^{ut,k}$ is the ratio of flow of utility added in the mixing task to the inlet flow rate to this task, $\lambda_2^{ut,k}$ corresponds to the ratio of utility added to post-mixing flow rate, and $\lambda_3^{ut,k}$ is the ratio of utility to post-waste flow rate.

5.2.2 Block II: Superstructure model

The superstructure model consists of a flow model and a logic model. The flow model represents the flow of material across the network of intervals. The logic model excludes unfeasible combinations of intervals from the superstructure and activates or de-activates the value of continuous variables using binary variables [66].

Flow model

Stream mixing task

The component inlet flow rate to the mixing task of an interval *k*, $f_{IN}^{i,k}$, is calculated as

$$f_{IN}^{i,k} = \sum_{kk} f^{i,kk,k} \quad \forall i, k \tag{5.9}$$

where $f^{i,kk,k}$ is the component flow rate of a stream from any interval *kk* to interval *k*.

Stream splitting task

The mass balance around the stream divider of the primary stream at the outlet of each interval is expressed as

$$f_{OUT,1}^{i,k} = \sum_{kk} f_1^{i,k,kk} \quad \forall i, k \notin PROD(k) \tag{5.10}$$

where $f_{OUT,1}^{i,k}$ is the inlet flow rate and $f_1^{i,k,kk}$ represent the flow rate of the outlets of the primary stream divider.

The mass balance around the stream divider of the seconday stream at the outlet of each interval is

$$f_{OUT,2}^{i,k} = \sum_{kk} f_2^{i,k,kk} \quad \forall i, k \notin PROD(k) \tag{5.11}$$

where $f_{OUT,2}^{i,k}$ are the inlet flow rates and $f_2^{i,k,kk}$ represent the flow rate of the outlets of the secondary stream divider.

Overall interval mass balance

An overall mass balance of the process interval to ensure mass balance consistency is formulated as

$$f_W^{i,k} = \sum_{kk} f^{i,k,kk} \quad \forall i,k \in PROD(k) \tag{5.12}$$

Equation (5.12) ensures that the total outlet flow rate of an interval corresponds to the sum of flow rates of streams connecting the interval to other intervals in the network.

Superstructure connections

Primary superstructure connections are defined as

$$f_1^{i,k,kk} \leq \zeta_P^{k,kk} f_{OUT,1}^{i,k} \quad \forall i,k,kk \tag{5.13}$$

where $\zeta_P^{k,kk}$ is a binary parameter representing the existence of a primary connection between intervals k and kk.

Similarly, secondary superstructure connections satisfy

$$f_2^{i,k,kk} \leq \left(\zeta^{k,kk} - \zeta_P^{k,kk}\right) f_{OUT,2}^{i,k} \quad \forall i,k,kk \tag{5.14}$$

where $\zeta^{k,kk}$ is a binary parameter representing the existence of a connection (primary or secondary) between intervals k and kk.

Note that, when recycles are present, the model remains linear if both the separation leading to a recycle stream and the purge have fixed, known parameters.

Logic model

Activation and deactivation constraints

Activation and deactivation constraints are added to activate or de-activate continuous variables based on the values of binary variables [66]. Take, for instance, an internal flow rate in an interval of the superstructure, $f^{i,k}$. This variable should be zero when the process unit is not selected or does not exist (that is, when $y^k = 0$), while if the interval is selected (*i.e.*, $y^k = 1$), then $f^{i,k}$ should take any value between an upper and a lower bound. This is expressed as

$$f_{LO}^{i,k} y^k \leq f^{i,k} \leq f_{UP}^{i,k} y^k \tag{5.15}$$

where $f_{LO}^{i,k}$, $f_{UP}^{i,k}$ are the lower and upper bounds of the flow rate $f^{i,k}$.

In this model, this type of constraints are used to activate and de-activate process flow rates, which are all defined as positive variables and are of the same order of magnitude.

Therefore, lower bound constraints are already enforced and do not need to be included. Moreover, a single upper bound, M, is used for most flow rates in the superstructure. The following activation constraints are defined:

$$f_W^{i,k} \leq y^k M \quad \forall i,k \tag{5.16}$$

which represents the activation/de-activation of the internal flow rates of an interval,

$$g_M^{i,k} \leq y^k M \quad \forall i,k \tag{5.17}$$

which activates/de-activates the flow rate of added chemicals to an interval,

$$\sum_i f_{IN}^{i,k} \leq y^k M \quad \forall k \tag{5.18}$$

which allows an interval to take inlet streams if and only if the interval exists in the selected structure, and

$$f^{i,k,kk} \leq \zeta^{k,kk} f_W^{i,k} \quad \forall i,k,kk \tag{5.19}$$

which enforces an upper bound on the interval-to-interval flow rates only if these are connected (that is, the connecting flow rate must be zero if there are no connections between intervals).

Logic expressions

Logic expressions are a set of conditions that avoid the selection of unfeasible solutions in the superstructure. Generally, these involve combinations of processing intervals that must or cannot be included in a solution at the same time. These are specified for each problem based on engineering knowledge about the alternatives being considered. In some cases, these can be enforced directly by using structural connections in the superstructure.

5.2.3 Block III: Location model

The location model is comprised of two sub-blocks: the calculation of location-based flow rates, and the location model.

Location-based flow model

The calculation of local flow rates is performed by taking the superstructure flow rates as a basis. For each interval, internal flow rates are converted into location-based flow rates. For any flow rate $f^{i,k}$, the corresponding location-based flow rate, $f_{LOC}^{i,k,l}$, satisfies

$$\sum_l f_{LOC}^{i,k,l} = f^{i,k} \quad \forall i,k \tag{5.20}$$

Location-dependent flow rates of an interval k in a location l, $f_{LOC}^{i,k,l}$, can only take a value when the section t containing the interval is allocated in location l, as given by

$$f_{LOC}^{i,k,l} \leq \sum_{t} \left(x^{t,l} \psi^{k,t} M \right) \quad \forall i,k,l \tag{5.21}$$

where $x^{t,l}$ is a binary variable representing the location of section t in location l, and $\psi^{k,t}$ is a 0-1 parameter indicating that interval k belongs to section t.

Equations (5.20) and (5.21) are implemented for the following location-based flow rates: post-waste flow rate, $f_{W,LOC}^{i,k,l}$, flow rate of added chemicals $g_{M,LOC}^{i,k,l}$, waste flow rate $g_{W,LOC}^{i,k,l}$, and flow rate of added utilities $g_{UTIL,LOC}^{ut,k}$, as described by Equations (5.22) to (5.29).

$$\sum_{l} f_{W,LOC}^{i,k,l} = f_{W}^{i,k} \quad \forall i,k \tag{5.22}$$

$$f_{W,LOC}^{i,k,l} \leq \sum_{t} \left(x^{t,l} \psi^{k,t} M \right) \quad \forall i,k,l \tag{5.23}$$

$$\sum_{l} g_{M,LOC}^{i,k,l} = g_{M}^{i,k} \quad \forall i,k \tag{5.24}$$

$$g_{M,LOC}^{i,k,l} \leq \sum_{t} \left(x^{t,l} \psi^{k,t} M \right) \quad \forall i,k,l \tag{5.25}$$

$$\sum_{l} g_{W,LOC}^{i,k,l} = g_{W}^{i,k} \quad \forall i,k \tag{5.26}$$

$$g_{W,LOC}^{i,k,l} \leq \sum_{t} \left(x^{t,l} \psi^{k,t} M \right) \quad \forall i,k,l \tag{5.27}$$

$$\sum_{l} g_{UTIL,LOC}^{ut,k,l} = g_{UTIL}^{ut,k} \quad \forall i,k \tag{5.28}$$

$$g_{UTIL,LOC}^{ut,k,l} \leq \sum_{t} \left(x^{t,l} \psi^{k,t} M \right) \quad \forall i,k,l \tag{5.29}$$

Constraints are included for the flow rates of added chemicals and waste in feedstock intervals,

$$g_{M,LOC}^{i,k,l} = 0 \quad \forall i,k \in RAW\,(k)\,,l \tag{5.30}$$

$$g_{W,LOC}^{i,k,l} = 0 \quad \forall i,k \in RAW\,(k)\,,l \tag{5.31}$$

which is done since these flows have no physical meaning and should remain inactive.

Inlet and outlet model

The inlet-outlet model defines specifications and constraints regarding inlet and outlet intervals of the superstructure, that is, raw material and product intervals. This model includes the calculation of total location-based total flow rates, the specification of feedstock compositions, the plant capacity specification (in terms of feedstock or product), and supply/demand constraints.

Total inlet/outlet flow rates

The total location-based flow rate of raw material, $F_{RAW}^{k,l}$, is calculated as

$$F_{RAW}^{k,l} = \sum_{i} f_{W,LOC}^{i,k,l} \quad \forall k \in RAW\,(k)\,, l \tag{5.32}$$

which is only applied to raw material intervals, $RAW\,(k)$, and the total location-based flow rate of product, $F_{PROD}^{k,l}$, is obtained as

$$F_{PROD}^{k,l} = \sum_{i} f_{W,LOC}^{i,k,l} \quad \forall k \in PROD\,(k)\,, l \tag{5.33}$$

which applies only to product intervals, $PROD\,(k)$.

Feedstock composition

The location-based biomass feedstock compositions are specified as

$$f_{W,LOC}^{i,k,l} = \phi^{i,k,l} F_{RAW}^{k,l} \quad \forall i,k \in RAW\,(k)\,, l \tag{5.34}$$

where $\phi^{i,k,l}$ is the composition of component i in feedstock k in location l.

Capacity specification

Capacity can either be specified in terms of flow rate of products, flow rate of feedstocks or only by variable bounds.

Either the rates of raw material or products rate can be specified. To specify the raw material rate

$$\sum_{l} F_{RAW}^{k,l} = FeedRate^{k,l} \quad \forall k \in RAW\,(k) \tag{5.35}$$

where $FeedRate^{k,l}$ is the given feedstock rate, and to specify the products rate

$$\sum_{p} F_{PROD}^{k,l} = ProdRate^{k,l} \quad \forall k \in PROD\,(k) \tag{5.36}$$

where $ProdRate^{k,l}$ is the given product rate.

Availability/demand constraints

An upper bound for the flow rate of feedstock from each location is defined as

$$\sum_{i} f_{W,LOC}^{i,k,l} \leq A^{k,l} \quad \forall k \in RAW\,(k)\,, l \tag{5.37}$$

where $A^{k,l}$ is the availability of feedstock k in location l.

The amount of product sold in a given market is bounded as

$$\sum_i f_{W,LOC}^{i,k,l} \leq D^{k,l} \quad \forall k \in PROD\,(k)\,, l \tag{5.38}$$

where $D^{k,l}$ is the demand of product k in location l.

Location-selection constraints

Location constraints are targeted towards processing intervals different from feedstocks and products, that is intervals representing units or processes. As shown in Figure 4.3 on page 41 and described in Section 4.4, the concept of processing sections is used in the location model. Processing sections are sets of processing intervals to be allocated together. This concept is used with the objective of creating a smaller mathematical problem size. Every location is represented by a binary variable, therefore allocating sections yields a lower number of binary variables than the number that would be obtained by allocating all intervals independently and enforcing the same location by adding additional constraints. In the worst case scenario, the same number of variables is obtained (*i.e.* if each single interval is to be allocated separately, leading to one section per interval).

Each section can only be placed in a location, which translates to the summation of locations in which a section is allocated being maximum 1, expressed as

$$\sum_l x^{t,l} \leq 1 \quad \forall t : t = 2, ..., N_{SEC} - 1 \tag{5.39}$$

where $x^{t,l}$ is a binary variable representing the location of section t in interval k. The only case in which a section is not allocated to any location is when none of the intervals in that section are selected.

Note that Equation (5.39) is applied to all sections except the first and last (feedstock and products). Different location selection rules are used for raw materials and products, given that the same raw material can be obtained from sources in different locations and a given product can be sold to multiple markets, therefore enforcing a single location for these intervals is not desired.

5.2.4 Block IV: Transportation model

This model block calculates the flow rate of material between geographical locations, $F_{LOC}^{l,l}$, which is required as input to the transportation cost model (see Section 5.2.5 on page 60).

First, component flow rates are added to calculate total flow rates between intervals

$$F_{INT}^{k,kk} = \sum_i f^{i,k,kk} \quad \forall k, kk \tag{5.40}$$

where $F_{INT}^{k,kk}$ is the total flow rate from interval k to interval kk.

The total flow rate between section t and section tt, $F_{SEC}^{t,tt}$, is computed as

$$F_{SEC}^{t,tt} = \sum_{k,kk} \left(F_{INT}^{k,kk} \psi^{k,t} \psi^{kk,tt} \right) \quad \forall t, tt \tag{5.41}$$

where $\psi^{k,t}$ is a 0-1 parameter representing the location of interval k in section t.

The flow rate between locations l and ll is obtained from

$$F_{LOC}^{l,ll} = \sum_{t,tt} \left(F_{SEC}^{t,tt} x^{t,l} x^{tt,ll} \right) \quad \forall l, ll \tag{5.42}$$

where $x^{t,l}$ is a binary variable representing the location of section t in location l. Therefore, these form a nonlinear product that is linearized in Section 5.4.2 to maintain the problem linearity.

5.2.5 Block V: Economic models

Capital cost model

The capital cost of the processing network, C_{CAP}, is computed as

$$C_{CAP} = \sum_{k} c_{EQ}^{k} \tag{5.43}$$

where c_{EQ}^{k} is the equipment cost of interval k.

The capital cost of processing intervals, c_{EQ}^{k}, is calculated using any expression of the equipment cost as a function of the capacity of each processing interval (in terms of flow rate). This expression might vary across different intervals and can therefore be selected individually for each interval. An example of a widely used equation for the equipment cost is the power law function

$$c_{EQ}^{k} = \alpha^{k} {F_{CAP}^{k}}^{\beta^{k}} \tag{5.44}$$

where α^{k} and β^{k} are the cost function parameters for interval k.

The total flowrate in each interval is calculated from one of the intermediate flow rates as

$$F_{CAP}^{k} = \sum_{i} f_{M}^{i,k} \quad \forall k \tag{5.45}$$

where F_{CAP}^{k} is the total flowrate in interval k for capital cost calculations.

This is a nonlinear concave function, that is linearized in Section 5.4.1 to improve the efficiency of the solution algorithms as well as ensuring global optimality of the solution [66].

Gross revenue

The economic revenue is represented by the sale of product, yet additional revenue sources could be considered and included, for example credit for sale of generated electricity.

Product sales The revenue from products sales, S_{PROD}, is expressed as

$$S_{PROD} = \sum_{i,k,l} \left(P_{PROD}^{k,l} f_{W,LOC}^{i,k,l} \right) \tag{5.46}$$

where $P_{PROD}^{k,l}$ is the location-dependent price of product k.

Operating cost model

The operating cost model consists of several terms that represent the contributions of different items.

Cost of raw materials The cost of raw material or feedstocks purchase, C_{RAW}, is

$$C_{RAW} = \sum_{i,k,l} \left(P_{RAW}^{k,l} \left(1 - \vartheta^{k,l} \right) f_{W,LOC}^{i,k,l} \right) \tag{5.47}$$

where $P_{RAW}^{k,l}$ is the location-dependent price of feedstock k and $\vartheta^{k,l}$ is the moisture content of raw material k in location l. This is done since, generally, biomass prices are given on a dry-weight basis, however, if that is not the case, then moisture can be set to 0, hence having no effect.

Cost of added chemicals The cost of added chemicals, C_{CHEM}, is computed in a similar fashion

$$C_{CHEM} = \sum_{i,k,l} \left(P_{CHEM}^{i,l} g_{M,LOC}^{i,k,l} \right) \tag{5.48}$$

where $P_{CHEM}^{i,l}$ is the location-dependent price of chemical i.

Cost of utilities consumed The cost of utilities, C_{UTIL}, is

$$C_{UTIL} = \sum_{ut,k,l} \left(P_{UTIL}^{ut,l} g_{U,LOC}^{ut,k,l} \right) \tag{5.49}$$

where $P_{UTIL}^{i,l}$ is the location-dependent price of chemical i.

Waste handling penalty The cost of waste handling, C_{WASTE}, is

$$C_{WASTE} = \sum_{i,k} P_{WASTE} g_{W}^{i,k} \tag{5.50}$$

where P_W is the price of waste handling per unit mass.

Transportation cost model The cost of transportation between locations,C_{TRANS} , is modeled as follows

$$C_{TRANS} = \sum_{l,ll} \left(\eta^{l,ll} P_{TRANS} F_{LOC}^{l,ll} \right) \tag{5.51}$$

where $\eta^{l,ll}$ is the distance between locations l and ll, P_{TRANS} is the transportation price, and $F_{LOC}^{l,ll}$ is the flow rate .

5.2.6 Block VI: Objective function

A variety of objective functions can be used as long as they require input data that is calculated in this model, otherwise models might need to be added to calculate inputs to new objective functions. In this thesis, an economic performance indicator, z, is used as objective function. This can take various forms, two common indicators are Gross Operating Income (GOI) and the Earning Before Interests and Tax (EBIT). The GOI is defined as

$$z_{GOI} = S_{PROD} - C_{RAW} - C_{CHEM} - C_{UTIL} - C_{TRANS} - C_{WASTE} \tag{5.52}$$

where S_{PROD} is the revenue from product sales, C_{RAW} is the cost of raw materials, C_{CHEM} is the cost of added chemicals (*e.g.*, solvents, co-reactants, C_{UTIL} is the utilities cost, C_{TRANS} is the transportation cost, and C_{WASTE} is the waste handling cost.

The second economic indicator, EBIT, is expressed as

$$z_{EBIT} = S_{PROD} - C_{RAW} - C_{CHEM} - C_{UTIL} - C_{TRANS} - C_{WASTE} - \frac{C_{CAP}}{\tau} \tag{5.53}$$

where S_{PROD} is the revenue from product sales, C_{RAW} represents the cost of raw materials, C_{CHEM} denotes the cost of chemicals added into the process, C_{UTIL} is the cost of utilities, C_{TRANS} is the transportation cost, and C_{CAP} denotes the capital cost, which is distributed into τ total time periods in the project lifetime.

5.3 Model generation

An appropriate model for a given synthesis problem can be generated from combinations of the model blocks presented in Sections 5.2.1 and 5.2.6. Two main models are used in the framework described in this thesis, both constituted by the presented model blocks, they are: (i) a *basic* mathematical model, and (ii) an *extended* mathematical model. The generation of these models is defined in Table 5.1.

Generation of desired models is done by combining model blocks. The final form of the optimization problem depends on the nature of each block, that is, if a set of nonlinear equations is included, the problem becomes nonlinear. The linear or nonlinear nature

Table 5.1. Definition of model blocks in terms of their equations, linearity (L is linear and NL is non linear) and guidelines for generation of the basic model for synthesis and the extended model for location-dependent synthesis from the defined model blocks.

Model block	Equations	L/NL	Basic model	Extended model
Processing interval	(5.1) to (5.8)	L	x	x
Superstructure	(5.9) to (5.14) and (5.16) to (5.19)	L	x	x
Location	(5.22) to (5.39)	L	–	x
Transportation	(5.40) to (5.42)	NL	–	x
Capital cost	(5.43) and (5.44)	NL	x	x
Operating cost	(5.46) to (5.50)	L	x	x
Transportation cost	(5.51)	L	–	x
Objective function	(5.52) or (5.53)	L	x	x

of the model formulations of each model block are specified in Table 5.1. Note that a model block is considered nonlinear if at least one of the equations in it is nonlinear.

5.4 Solution strategies

Solution strategies are used to reformulate the optimization problem in such a way that the solution is reached more easily and/or in a more reliable manner. Moreover, a strategy to generate not only a single optimal solution, but a set of near-optimal solutions is discussed.

Optimization problems with nonlinear model formulations can be solved through the use of appropriate solution algorithms (see 5.6). However, the solution of nonlinear poses certain difficulties in terms of the efficiency of solution algorithms and the existence of nonconvexities, that may lead to local optimal solutions, rather than the global optimum [66].

In this section, linearization approaches for different types of nonlinearities in the presented model are described.

5.4.1 Linearization of continuous nonlinear functions

Continuous nonlinear terms may be present in synthesis models, thus introducing a source of nonlinearities. In the presented model, the capital cost model as a function

of the throughput in each interval represents a nonlinear concave function, which can be approximated using piecewise underestimators [73, 66, 74].

Piece-wise linearization of the capital cost model

A piece-wise linearization of the capital cost model is shown in this section.

Piecewise underestimators are formulated in disjunctive form [74] as follows.

Equipment cost of each interval The linear approximation of $c_{EQ}\left(F_{CAP}^{k}\right)$, Equation (5.44), is expressed for each piece j as

$$c_{EQ}^{k} = \sum_{j<J} \left(\alpha_{LIN}^{j,k} F_{D}^{j,k} + \beta_{LIN}^{j,k} w^{j,k} \right) \quad \forall k \tag{5.54}$$

where $\alpha_{LIN}^{j,k}$ and $\beta_{LIN}^{j,k}$ are the parameters of each underestimator j in interval k and $w^{j,k}$ is a binary variable representing the "selection" of piece j.

Selection of maximum one disjunctive term per interval The variable $w^{j,k}$ takes a value of 1 if the total flow rate of the considered interval is within the range of piece j. Only one disjunctive term can be selected per interval, as expressed by

$$\sum_{j<J} w^{j,k} = 1 \quad \forall k \tag{5.55}$$

where $w^{j,k}$ are binary variables representing the selection of piece j in interval k. In this context, the term piece refers to a range of flowrates, since the flowrate domain (independent variable) is partitioned into multiple "pieces".

Disaggregated flow rate in each interval The summation of disaggregated flows rates should correspond to the total flow rate, as expressed by

$$F_{CAP}^{k} = \sum_{j}^{J-1} F_{D}^{j,k} \quad \forall k \tag{5.56}$$

where $F_{D}^{j,k}$ is the piece j disaggregated flow rate in interval k. Moreover, two constraints are added so that the disjunctive terms are properly activated or deactivated, as

$$\Gamma^{j} w^{j,k} \leq F_{D}^{j,k} \leq \Gamma^{j+1} w^{j,k} \quad \forall k, j < J \tag{5.57}$$

where Γ^{j} and Γ^{j+1} are the flow rate bounds of piece j.

Equations (5.54) to (5.57) can be used to replace Equation (5.44) in the original model, so as to obtain a linear formulation.

5.4.2 Linearization of bilinear products

Bilinear products of binary–continuous and binary–binary variables appear in the transportation model block (see 5.2.4). These products are linearized using the approach outlined by Floudas [66].

Linearization of transportation model

The linearization of Equation (5.42) is considered below, which contains a product of a continuous variables and two binary variables. The linearization is thus done in two sequential steps: first, the bilinear product of binary variables is linearized, and then, the resulting bilinear continuous–binary product is addressed.

As mentioned, the nonlinear product $F_{SEC}^{t,tt}x^{t,l}x^{tt,ll}$ is linearized in two steps. First, the bilinear product of two binary variables $x^{t,l}x^{tt,ll}$ is linearized, by substituting it with a new variable $TR^{t,tt,l,ll}$ and adding the necessary constraints. Next, the resulting bilinear product $F_{SEC}^{t,tt}TR^{t,tt,l,ll}$ of a continuous and a 0-1 variable is linearized, by introducing a new variable $h^{t,tt,l,ll}$ and a set of constraints.

The linearization of the bilinear product $x^{t,l}x^{tt,ll}$ is done by introducing new variables $TR^{t,tt,l,ll}$ such that $TR^{t,tt,l,ll} = x^{t,l}x^{tt,ll}$ and introducing three additional constraints for each (t,tt,l,ll)

$$TR^{t,tt,l,ll} \leq x^{t,l} \quad \forall t,tt,l,ll \tag{5.58}$$

$$TR^{t,tt,l,ll} \leq x^{tt,ll} \quad \forall t,tt,l,ll \tag{5.59}$$

$$TR^{t,tt,l,ll} \geq x^{t,l} + x^{tt,ll} - 1 \quad \forall t,tt,l,ll \tag{5.60}$$

The product $F_{SEC}^{t,tt}x^{t,l}x^{tt,ll}$ becomes $F_{SEC}^{t,tt}TR^{t,tt,l,ll}$, which is linearized by introducing a variable $h^{t,tt,l,ll}$ so that $h^{t,tt,l,ll} = F_{SEC}^{t,tt}TR^{t,tt,l,ll}$ and introducing four additional constraints for each (t,tt,l,ll)

$$F_{SEC}^{t,tt} - M\left(1 - TR^{t,tt,l,ll}\right) \leq h^{t,tt,l,ll} \leq F_{SEC}^{t,tt} \tag{5.61}$$

$$0 \leq h^{t,tt,l,ll} \leq TR^{t,tt,l,ll}M \tag{5.62}$$

Since $h^{t,tt,l,ll}$ is already defined as a positive continuous variable, the non-negativity constraint is already enforced and does not need to be added. As a result, only three constraints are required in the model.

The set of Equations (5.58) to (5.59) may be introduced replacing Equation (5.42) in order to obtain a linear formulation.

5.4.3 Generation of top-ranked solutions

Integer cuts are implemented so as to generate a ranking of solutions [75].

These can be introduced on the selection of processing intervals (network topology), as

$$\sum_{k} \left(2Y^{k,c} - 1\right) y^{k} \leq \sum_{k} Y^{k,c} \quad \forall c \tag{5.63}$$

where y^k are binary variables indicating the selection of intervals in the solution, and $Y^{k,c}$ is a parameter that stores the solution for each cut c.

Alternatively, integer cuts can be introduced on the location-selection for processing sections, as

$$\sum_{t,l} \left(2X^{t,l,c} - 1\right) x^{t,l} \leq \sum_{t,l} X^{t,l,c} \quad \forall c \tag{5.64}$$

where $x^{t,l}$ are binary variables indicating the selection of a location for a section, and $X^{t,l,c}$ is a parameter that stores the solution for each cut c.

The implementation of integer cuts is done using a loop that runs on each cut, solving the problem while excluding previous solutions and storing each additional solution.

5.5 Input data

Multiple data parameters are required as input to the model. These are listed in Table 5.2 by model block, as depending on the blocks included in the model, the appropriate parameters need to be provided.

The combinatorial nature of the synthesis problem may originate a very large number of alternatives to be considered, which given the number of parameters involved in the models, even for simpler models, can result in a large amount of data values. For this reason, data management approaches should be included in process synthesis methods, especially for specific domains where data is scattered and not systematized, such as the design of biorefineries. This issue is addressed in Chapter 7.

5.6 Optimization methods

The mathematical form of the synthesis problem represented by the equations in section Section 5.1 and the resulting optimization problem is to be solved in order to find optimal processing routes.

The synthesis problem is a decision-making problem, which by nature follows the form of a Mixed-Integer Nonlinear Program (MINLP). However, it can, under certain conditions (as shown in Section 5.4), be approximated to a Mixed-Integer Linear Program

Table 5.2. Model parameters required by each model block.

Model block	Sub-block	Model parameters by model block
Processing interval	Mixing	$\mu^{i,ii,k}$
	Reaction	$\theta^{reac,k,r}$, $\gamma^{i,r}$
	Waste separation	$\delta^{i,k}$
	Product separation	$\sigma^{i,k}$
	Utility consumption	$\lambda_1^{ut,k}$, $\lambda_2^{ut,k}$, $\lambda_3^{ut,k}$
Superstructure	Connections	$\upsilon^{k,st}$, $\zeta^{k,kk}$, $\zeta_P^{k,kk}$
	Activation	M
Location	Local flow rates	$\psi^{k,t}$
	Inlet nodes	$\phi^{i,k,l}$
	Availability/demand	$Av^{k,l}$, $Dem^{k,l}$
Transportation		$\psi^{k,t}$
Economic	Capital cost	$(\alpha^k, \beta^k) \vee (\alpha_{LIN}^{j,k}, \beta_{LIN}^{j,k}, \Gamma^j)$
	Operating cost	$P_{PROD}^{k,l}$, $P_{RAW}^{k,l}$, $P_{CHEM}^{i,l}$, $P_{UTIL}^{ut,l}$, P_W
	Transport cost	P_{TRANS}
Objective function	Lifetime	τ

(MILP or MIP) by using linear models for technologies for fixed ranges of operating conditions and/or through linear approximations of the nonlinear equations involved.

Appropriate mathematical solution algorithms for the optimization problem are selected based on the problem formulation (MILP or MINLP). In MILP problems, difficulties arise from the combinatorial nature of the 0-1 variables, which leads to a very large number of alternative solutions [66]. Algorithmic approaches for MILP problems include Branch and bound methods [76]. For MINLP problems, the main challenges are related to the size of the combinatorial problem and the nonlinearities, which can be nonconvex, hence leading to the existence of multiple local solutions [66]. Convex problems can be approached with algorithms such as Outer Approximation with Equality Relaxation (OA/ER) [77], which is implemented as DICOPT. For nonconvex problems, algorithmic developments based on the branch-and-bound concept have been made towards achieving global solutions and are available as BARON [21] and ANTIGONE [22].

5.7 Conclusion

The modeling approach of the framework has been described in this chapter. Generic model blocks are developed to model the superstructure of alternatives. Different models can be generated from combinations of said modeling blocks, depending on each problem needs. Solution strategies are outlined to cope with nonlinearities and for generating sets of near-optimal solutions, which may not be optimal but still have engineering value. Input parameters need to be provided, which allow the representation of specific problems using the generic models. Last, optimization methods are outlined.

CHAPTER 6

Workflow & tools

The detailed workflow and data flow of the biorefinery synthesis framework are described in this chapter along with an overview of computer-aided tools supporting the framework. Tools include an implementation of the framework as the software interface Super-O. The systematic workflow and data-flow allow a comprehensive execution of the required steps. Super-O allows non-expert users to execute the steps in the workflow and solve synthesis problems in a shorter time frame without prior knowledge of the modeling language or optimization techniques.

6.1 Workflow

The workflow for synthesis of biorefinery networks follows eight steps:

Step 1. Problem definition
Step 2. Alternatives definition
Step 3. Modeling of single-location problem
Step 4. Single-location solution
Step 5. Location alternatives definition
Step 6. Modeling of multi-location problem
Step 7. Multi-location solution
Step 8. Results reporting

A schematic representation of the workflow, data flow and integration with supplementary methods and tools is shown in Figure 6.1. Descriptions of each step in the workflow are presented in the following sections, which are given in terms of a general overview including the step objective, a list of actions to be performed (workflow), data inputs and outputs (data flow) and a list of the necessary methods and tools.

The workflow consists of a first step for the overall problem definition, followed by steps 2-4 towards the solution of the single-location problem. Next, the multi-location dependent problem is defined and solved in steps 5-7. Finally, results are summarized in step 8. A summary of the workflow actions is given in Table 6.1.

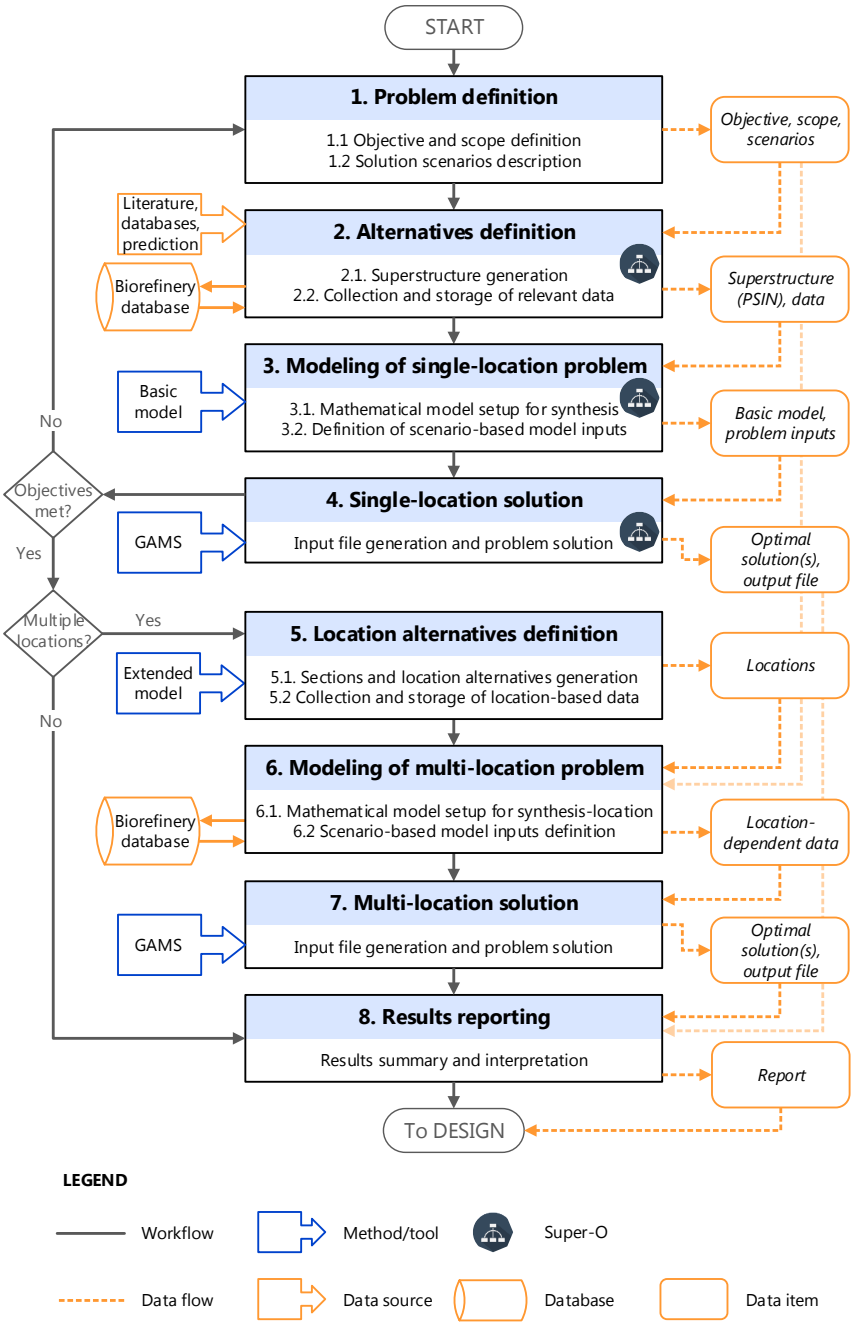


Figure 6.1. Workflow and data flow for the biorefinery synthesis framework.

Table 6.1. Summary of the workflow for biorefinery synthesis.

Step	Sub-step	Action	Description
1	1.1	1.1.1	State objective
		1.1.2	Define scope
		1.1.3	List additional performance indicators
	1.2	1.2.1	Specify location issues
		1.2.2	Define scenarios
		1.2.3	Select single solution or set of top ranked solutions
2	2.1	2.1.1	Perform database search
		2.1.2	Generate/expand superstructure
	2.2	2.2.1	Retrieve data from database
		2.2.2	Collect data
		2.2.3	Estimate missing data
		2.2.4	Input data in Super-O or input file
		2.2.5	Store new data in database
3	3.1	3.1.1	Select/define objective function
		3.1.2	Select model blocks
		3.1.3	Add additional constraints
		3.1.4	Include or not integer cut constraints
	3.2	3.2.1	Define scenario-dependent model inputs
		3.2.2	Add inputs to Super-O or update input file(s)
4	4.1	4.1.1	Select solution algorithm
		4.1.2	Solve for each scenario
		4.1.3	Continue to Step 5 or skip to Step 8
5	5.1	5.1.1	Define set of locations
		5.1.2	Define processing sections
	5.2	5.2.1	Perform database search
		5.2.2	Collect data
		5.2.3	Input data in Super-O or input file
		5.2.4	Store new data in database
6	6.1	6.1.1	Select objective function
		6.1.2	Select model blocks
		6.1.3	Add additional constraints
		6.1.4	Include or not integer cut constraints
	6.2	6.2.1	Define scenario-dependent model inputs
		6.2.2	Add inputs to Super-O or update input file(s)
7	7.1	7.1.1	Select solution algorithm
		7.1.2	Solve for each location-dependent scenario
8	8.1	8.1.1	Store problem definition
		8.1.2	Generate report of scenarios
		8.1.3	Store model information
		8.1.4	Store input file information
		8.1.5	Store output file information

6.1.1 Step 1: Problem definition

The objective of this step is to define the synthesis problem that needs to be solved by specifying its objective and scope. Moreover, various scenarios to be considered are specified.

The problem objective is to be translated into an objective function in Step 3. Other problem elements (feedstock-process-product) can be specified at this step, limiting the search space that is to be mapped in Step 2. Examples of these are a (set of) raw materials to be considered, a set of geographical locations or a list of end products. Additional performance indicators that are to be calculated post-optimization are also defined in this step. Next, scenarios for the problem solution are defined.

Step 1.1: Objective and scope definition

Action 1.1.1: State problem objective in terms of performance criterion to minimize/-maximize.

Action 1.1.2: Define problem scope in terms of:

a) raw materials and/or products
b) type of technologies allowed
c) geographical location(s)

Action 1.1.3: List additional performance indicators that are to be calculated post-optimization.

Output: Problem objective and scope, constraints for alternatives and performance indicators

Note 1: Specifications regarding regarding the problem scope in Action 1.1.2 can be done in terms of specific feedstocks/processing alternatives/products or by specifying the type of alternatives that are to be considered for the problem. For example, biochemical transformations can be chosen as the only type of transformations allowed for a given problem.

Step 1.2: Solution scenarios definition

Action 1.2.1: Specify if location and transportation issues need to be included in the solution (*i.e.*, if the process synthesis problem is of interest or both process and supply chain synthesis aspects are to be considered).

Action 1.2.2: For both problem types, define the scenarios under which each problem is to be solved with respect to:

a) Terms considered in the objective function
b) Values of parameters to vary

Action 1.2.3: For each scenario, state whether a single optimal solution or a set of top ranked solutions are to be obtained.

Input: Problem objective and scope

Output: Scenarios

6.1.2 Step 2: Alternatives definition

The objective of step 2 is to define the search space for the problem by postulating a superstructure of alternatives of interest and collecting all the necessary data for the mathematical solution of the problem. This step is divided into: (2.1) superstructure generation, and (2.2) data collection.

To fully specify the search space, the following are to be defined: (i) a set of available raw materials, (ii) a set of desired products, (iv) a set of processing steps, (v) and a set of available processing intervals to perform the tasks involved in each of the considered steps.

Initially, a database search is performed in order to identify possible alternatives for the problem given the elements defined in Step 1 as inputs of the search. A network of alternatives is generated through a database search by specifying one or a set of raw materials or products. Moreover, search constraints can be set up according to those specifications defined in Step 1. If a network has been found in the database containing relevant alternatives, this is taken as initial step towards reviewing and expanding the superstructure. Else, if a network has not been found during the database search, then the superstructure is generated from scratch.

In the data collection step, data for each of the alternatives in the superstructure, including raw materials, processing alternatives and products, is collected. If the alternatives are obtained from the database, data is already available and can be kept as is or modified. In case the alternatives are not obtained from the database, data needs to be collected and converted into the correct form for it to be taken as input by the generic model. Alternatives data is collected from various sources, including open literature, industrial or academic partners, and databases.

When data is not readily available, it may be predicted or estimated. Component property data can be predicted using property prediction tools such as ProPred [78]. Stoichiometric data for reactions, when not available, can be estimated using, for example, black box experimental data via calculating the maximum theoretical yield using redox

balance [79]. Data regarding the separation efficiency is estimated using thermodynamic calculations such as the driving force [80]. After data collection/prediction, the data is stored in the Biorefinery Synthesis Database for future reference and reuse.

Step 2.1: Superstructure generation

Action 2.1.1: Perform a database search in the Biorefinery Synthesis Database given the specifications defined in Step 1, that is, by specifying feedstocks/products and adding search constraints to the types of alternatives allowed. If available data is not to be considered for the problem, perform Action 2 directly.

Action 2.1.2: If a superstructure of alternatives is available from Action 1.1.1, take that as a starting point and expand to cover the desired search space. If a superstructure of alternatives has not been found in Action 1.1.1, then generate a superstructure from scratch.

Input: Problem definition

Output: PSIN superstructure of alternatives

Methods/tools: Biorefinery Synthesis Database, external data sources (literature, databases, etc.)

Note 1: Quaglia [71] classified superstructure generation methods as: (i) alternative collection, (ii) combinatorial synthesis, and (iii) insight-based synthesis. Recently, a computer-aided tool has been presented, that generates a superstructure using insights-based synthesis [17], which can be used in conjunction with this framework.

Step 2.2: Collection and storage of relevant data

Action 2.2.1: Retrieve alternatives data from database (see network generation algorithms in Section 7.4.1).

Action 2.2.2: or new alternatives, collect data from literature, databases, experiments or industrial partners.

Action 2.2.3: Estimate missing data using prediction methods.

Action 2.2.4: Input data in Super-O (non-expert user) or in a Super-O input file (expert user).

Action 2.2.5: Store any new data in the Biorefinery Synthesis Database.

Input: PSIN superstructure of alternatives

Output: Necessary data for all alternatives in input files(s)

Methods/tools: Biorefinery Synthesis Database, external data sources (literature, databases, etc.), Super-O

6.1.3 Step 3: Modeling of single-location problem

The objective of the modeling step is to formulate the mathematical model of the superstructure following the generic formulation in **??** on page ??. This step consists of: (3.1) model setup, and (3.2) definition of scenarios and model inputs.

The mathematical optimization problem is set up in this step using a combination of the model blocks presented in Chapter 5 and the selected performance indicator as objective function. Moreover, additional constraints may be added for each specific problem.

Once the model details are set up, scenarios for the problem solution are defined and model inputs for each scenario need to be defined. Model inputs to be defined in this step are: feedstock(s) or product(s) flow rate, number of integer cuts and values of parameters varied across scenarios. Input files for each scenario are generated, by adding the specific scenario data to the data collected in Step 2.

Step 3.1: Mathematical model setup for synthesis

Action 3.1.1: Select objective function based on the mathematical translation of the problem objective defined in Step 1.

Action 3.1.2: Select relevant model blocks to form the optimization model for the problem. The basic model or a variation of it is used in this step (see Figure 5.1).

Action 3.1.3: Add any necessary additional constraints. Examples of additional constraints are the selection of one interval per step, forcing combinations of intervals to be selected together, or specifying forbidden matches.

Action 3.1.4: Include integer cut constraints if ranked solutions are to be obtained, based on what has been specified in Step 1.

Input: PSIN superstructure of alternatives

Output: Mathematical model

Methods/tools: Biorefinery Synthesis Database, Super-O

Note 1: Note that the selection of an objective function different from those presented in 5.2.6 in any of its variations, may result in additional data being required, which is manually input.

Note 2: At this point the basic synthesis problem is considered (*i.e.*, considering a single location), hence location and transportation blocks of the model are not considered (see Table 5.1 on page 61).

Step 3.2: Definition of scenario-based model inputs

Action 3.2.1: Based on scenarios defined in Step 1.2, define scenario-dependent model inputs: flow rate of feedstock(s) or product(s), number of integer cuts and values of parameters varied across scenarios.

Action 3.2.2: Update input file(s) with scenario-dependent model inputs.

Input: PSIN superstructure and mathematical model

Output: Scenarios and updated input file(s)

Methods/tools: Super-O

Note 1: Scenarios defined in Step 1.2 are expressed in terms of variations in values of the parameters considered (*e.g.*, considering variations in the product market price), variations in the objective function (*e.g.*, and including/not including waste handling costs).

6.1.4 Step 4: Single-location solution

The optimization problem is solved in this step by employing solution algorithms implemented in GAMS [72]. The inputs to GAMS are the model file and the input file with all the necessary problem data (model parameters, raw material data, cost data, etc.). The outputs from the solver are the optimal values of the objective function, the corresponding optimization variables, and all other process variables, which are all given in an output file accessible through Super-O. The problem solution can be obtained through the user interface Super-O or directly in GAMS.

Step 4.1: Input file generation and problem solution

Action 4.1.1: Select solution algorithm based on the model formulation.

Action 4.1.2: Solve for each scenario.

Action 4.1.3: If a location-dependent problem has been defined in Action 1.2.1, then go to Step 5, else skip to Step 8.

Input: Mathematical model file and input file

Output: Output file with problem solution

Methods/tools: Super-O and GAMS

Note 1: The choice of solver is made based on the model formulation (see Table 6.2).

Table 6.2. Optimization algorithms for different problem forms available through GAMS [72].

Problem form	Algorithm	Solver	Ref.
Linear	Branch and cut / Simplex (Branch-and-bound)	CPLEX	[76]
Convex non-linear	Outer Approximation / Equality Relaxation (OA/ER)	DICOPT	[81]
Non-convex	Branch-and-reduce	BARON	[82, 21]
	Branch-and-bound global optimization	ANTIGONE	[22]

6.1.5 Step 5: Location alternatives definition

Multiple locations to be considered in a synthesis-location problem are defined in this step. Locations are defined based on the problem needs as geographical regions of any size. These constitute the set of geographical locations for the problem. Moreover, processing sections are defined at this point. The number of processing sections correspond to the number of parts in which the processing network is divided for location selection purposes. For each section, an optimal location will be determined.

Given the set of locations defined in the previous step, location-dependent data is collected in this step for all location-dependent parameters. The main parameters are market prices of feedstock, chemicals, utilities, products, availability of feedstocks, demand of products, feedstock composition, and distances between geographical locations.

Step 5.1: Sections and location alternatives generation

Action 5.1.1: Define set of locations as locations for feedstock harvesting/purchase, product markets and plant location(s).

Action 5.1.2: Define processing sections for location selection as sets of intervals. Rules for defining processing sections are:

a) At least one interval needs to be included in each processing section.
b) Each interval can only be allocated to one section.
c) Feedstock and product intervals need not be assigned to any processing sections, as being the first and last steps in the network, they are automatically identified as sources and sinks.

Input: PSIN superstructure of alternatives (without locations)

Output: Set of locations, processing sections, extended PSIN superstructure

Tools: Biorefinery Synthesis Database

Note 1: Processing sections are selected based on the physical boundaries of the subprocesses included in the network and defined as groups of processing intervals. It is important to note that the locations in the list defined at this point include those for location of processing sections as well as for feedstock purchase and sale of products.

Note 2: In terms of generating a list of relevant locations for the problem, various considerations need to be made. First the problem scale is determined as global or regional, which can be specified as world-wide, country-wide, or state-wide, among others. Then, locations within the considered area are to be defined by either considering political and well-defined areas or other custom-made boundaries based on weather conditions or biomass growth patterns, among many other options. In this regard, Geographic Information Systems (GIS) are gaining interest by the Process Systems Engineering (PSE) community for its ability to provide spatial information [83, 84].

Note 3: A database search in the Biorefinery Synthesis Database can be performed in order to identify relevant locations for given raw materials and products.

Note 4: Process sections are defined as sets of intervals. However, all intervals in a step are generally allocated in the same section. Therefore, one can think of process sections as sets of processing steps. However, there are problem types where this does not hold (see Section 4.6), hence defining sections in terms of intervals only is allowed.

Step 5.2: Collection and storage of location-based data

Action 5.2.1: Perform a database search in the Biorefinery Synthesis Database to retrieve location-dependent data for relevant items from the following:

a) Feedstocks availability, composition, price
b) Products demand, market price
c) Price of added chemicals
d) Price of utilities
e) Distance between geographical locations
f) Transportation price

Action 5.2.2: For data not available in the database, collect data from literature, databases or other sources.

Action 5.2.3: Input data into the input file.

Action 5.2.4: Store any new data in the Biorefinery Synthesis Database.

Input: ePSIN superstructure

Output: Location-dependent data, extended input file

Tools: Biorefinery Synthesis Database, literature, databases

Note 1: Note that data retrieved from the database can be updated/modified in this step for the given problem.

6.1.6 Step 6: Modeling of multi-location problem

The location-dependent problem is modeled using the extended model (see Section 5.3). Model blocks are selected in this step to form an extended optimization model for the location-dependent problem. Extended model blocks are related to location and transportation calculations.

Step 6.1: Mathematical model setup for synthesis-location

Action 6.1.1: Select objective function based on the mathematical translation of the problem objective defined in Step 1.

Action 6.1.2: Select relevant model blocks to form the optimization model for the problem. The extended model or a variation of it is used in this step (see Figure 5.1).

Action 6.1.2: Add any necessary additional constraints. Examples of these constraints are:

a) fixing the location of one of the processing sections
b) forbidden matches location-interval or interval-interval that have not been constrained in the ePSIN representation of the problem

Action 6.1.3: Include integer cut constraints if ranked solutions are to be obtained, based on the specifications of Step 1.

Input: Extended PSIN superstructure of alternatives (including locations)

Output: Location-dependent data and extended input file

Note 1: Integer cut constraints can be added on the selection of processing intervals (y^k), or on the selection of locations $x^{t,l}$. Thereby, a ranking of solutions in terms of processing route selection or in terms of location selections is obtained.

Step 6.2: Definition of scenario-based model inputs

Action 6.2.1: For location-dependent scenarios defined in Step 1.2, define scenario-dependent model inputs: flow rate of feedstock(s) or products, number of integer cuts and value of parameters varied across scenarios.

Action 6.2.2: Update input file(s) with scenario-dependent model inputs.

Input: Extended PSIN superstructure of alternatives (including locations) and mathematical model

Output: Location-dependent scenarios and updated input files

6.1.7 Step 7: Multi-location solution

The solution of the location-based problem is done through solvers implemented in GAMS [72] using the model and generated input file(s) as input and obtaining the corresponding output file(s).

Step 7.1: Input file generation and problem solution

Action 7.1.1: Select solution algorithm based on the model formulation.

Action 7.1.2: Solve for each location-dependent scenario.

Input: Extended mathematical model file and input file(s)

Output: Output file(s) with location-dependent problem solution

Methods/tools: GAMS

Note 1: The choice of solver is made based on the model formulation (see Table 6.2 on page 75).

6.1.8 Step 8: Summary report

This step entails the generation of a summary of the problem definition, scenarios, models, and generated input file(s) for each scenario and results extracted from the obtained output file(s).

Step 8.1: Generation of summary report

Action 8.1.1: Store problem definition.

Action 8.1.2: Generate report of scenarios.

Action 8.1.3: Store model information (selected model blocks and extra constraints added).

Action 8.1.4: Store input file information for each scenario.

Action 8.1.5: Store output (results) file information for each scenario.

Input: Output file(s) with basic and location-dependent problem solution

Output: Summary report

Methods/tools: EXCEL

6.2 Tools

A number of in-house and commercial tools are required in different steps of the presented workflow. These include property prediction tools, databases, optimization software and interfaces, which are summarized and described in Table 6.3.

Table 6.3. Description of tools used in the framework.

Name	Type	Description	Function	Step
ProCAMD[a]	In-house	Interface for Computer-Aided Molecular Design	Generation of reaction products	2
ProCARPS[b]	In-house	Interface for Computer-Aided Reaction Path Synthesis	Generation of reaction paths from a given feedstock to a given product	2
ProPred[c]	In-house	Interface for property prediction	Prediction of pure compound properties	2
Biorefinery Synthesis Database[d]	In-house	Synthesis data storage/retrieval of biorefinery synthesis data	Retrieval and storage of data	2, 6
Super-O[e]	In-house	Interface for formulation and solution of superstructure optimization problems	User guidance through the formulation and solution of synthesis problems	1, 2, 3, 4
EXCEL	External	Spreadsheets	Storage of input/output files data	4, 7
GAMS[f]	External	Modeling system for mathematical programming and optimization (language compiler and integrated solvers)	Solution of the optimization problem	4, 7

[a] Available in ICAS [30] [b] Cignitti [85] [c] Available in ICAS [30] [d] see Chapter 7
[e] see Section 6.2.1 [f] GAMS Development Corporation [72]

6.2.1 Software implementation: Super-O

The workflow highlighted in the previous section has been implemented in a software interface, which was initially named EOLO [20] and recently updated and re-named Super-O [86]. It has been developed in C# platform and it automates most of the tasks in steps 2, 3 and 4 of the framework.

The flow diagram of Super-O (Figure 6.2) highlights the different actions performed by the interface, as well as the interactions with other tools. Given data on a problem superstructure, the Super-O interface is able to structure the data, perform consistency checks on it, provide a graphical representation of the superstructure, calculate the parameters for the linearization of nonlinear functions (if needed), give access to the generic model to edit it, solve the optimization problem and open an output file containing the results. The items of the implemented workflow shown in Figure 6.2 are described below.

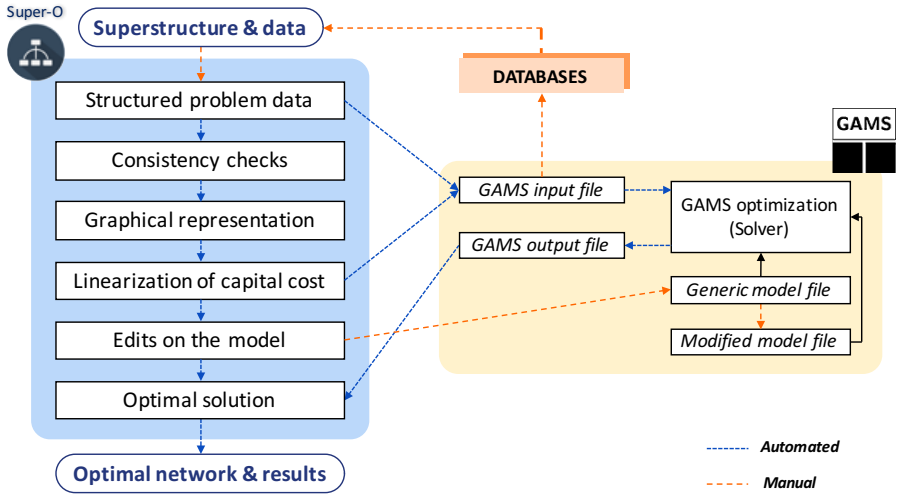


Figure 6.2. Flow diagram of Super-O, the user interface for formulating and solving synthesis problem using a superstructure optimization approach.

Structured problem data

Super-O assists non-expert and expert users in systematizing and structuring process synthesis data into the corresponding data layers. The problem size and structure are the first specifications input into the interface (see Figure 6.3), *i.e.*, number of processing steps, intervals, components, utilities, and reactions. From these, the appropriate number of data slots are generated for all the parameters that need to be input.

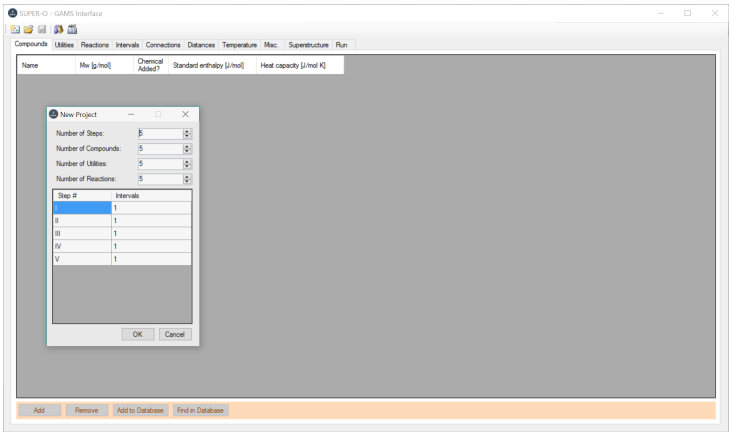


Figure 6.3. Super-O interface tab for general problem data input.

Consistency checks

Consistency checks are implemented in Super-O in order to ensure that the data represents a solvable problem as well as to minimize and identify possible errors when typing in data. For example, the consistency of stoichiometric data values is checked through a simple mass balance using molar weight values assigned to the components (see Figure 6.4).

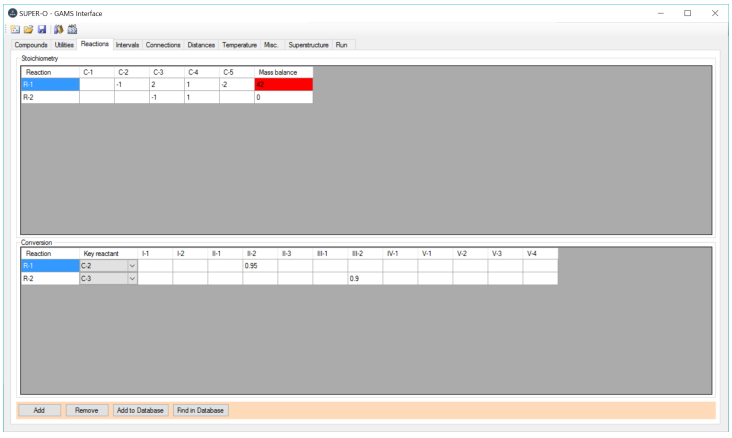


Figure 6.4. Super-O interface tab for reaction data input with consistency check.

Graphical representation

A graphical representation of the superstructure of alternatives for a given problem is generated and displayed by Super-O (see Figure 6.5). This provides a visualization of

the topology alternatives embedded in the superstructure. The user can visually review the generated superstructure and make changes if necessary.

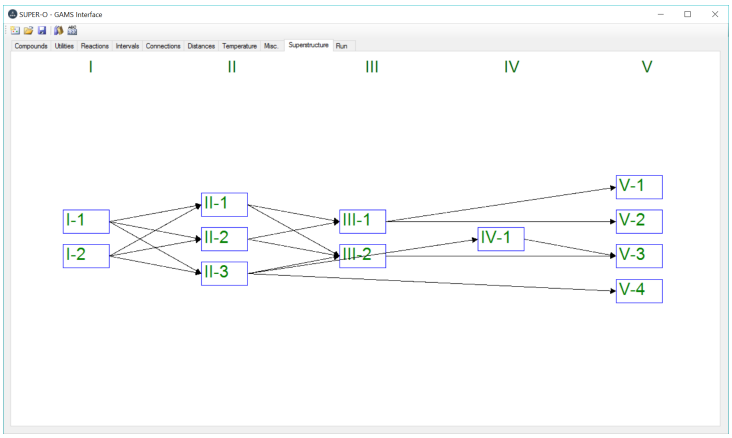


Figure 6.5. Super-O interface tab displaying the superstructure graphical representation.

Piecewise linearization of nonlinear functions

In cases where piecewise approximations of nonlinear functions for capital cost are used, parameters for the linear approximation need to be calculated for each piece. This is done automatically in Super-O for the equipment cost calculation of each interval (see Figure 6.6). The user inputs are: (i) a nonlinear function (expressed in Python syntax, see Table 6.4); (ii) a lower bound for the interval throughput (default value is 0); (iii) an upper bound for the interval throughput (default value is M); (iv) number of pieces for linearization.

Table 6.4. Examples of Super-O syntax for capital cost functions. Note that the variable f (`fpoint`) represents the throughput in each given interval.

Type	Mathematical expression	Super-O syntax
Linear	$20f$	`20*fpoint`
Power law	$100f^{0.6}$	`100*pow(fpoint,0.6)`

Edits on the model

Changes on the generated model can be input via Super-O, however, prior knowledge of the modeling language (GAMS) is required. Model modifications are mainly directed towards adding additional constraints, for example, constraints specifying forbidden matches or intervals that need to be selected together (structural constraints),

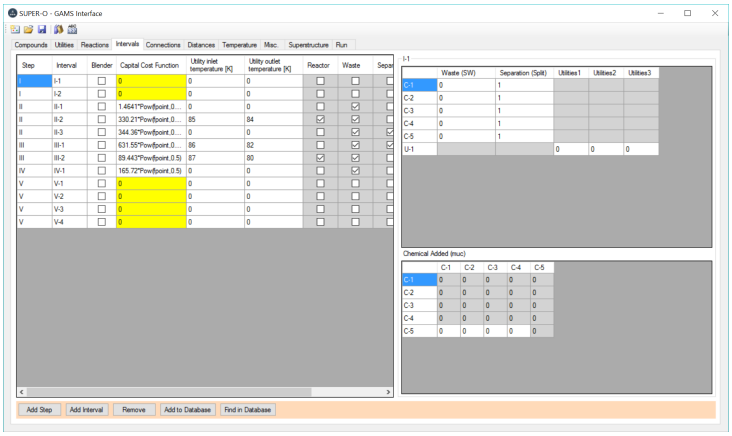


Figure 6.6. Super-O interface tab for intervals input data including capital cost functions specification.

constraints to fix certain variables for a given scenario (fixed value constraints), or extended models to represent phenomena that are not included in the processing interval model (process model constraints). Note that structural constraints are added only when these cannot be represented using connections in the superstructure or when this is not desired.

Optimal solution

A connection between Super-O and GAMS exists so that solvers implemented in GAMS are used to solve the optimization problem. The user has no interaction with GAMS since the solution report is shown in the Super-O screen and the model outputs are stored in an output file accessible from Super-O.

Tools integration

Based on the specified problem information, Super-O generates a GAMS-readable binary file containing all problem data for the solution of the optimization problem through the GAMS model file containing the generic model equations. Then, Super-O directly calls an optimization software, in this case GAMS [72], to solve the optimization problem and generate an output file containing the results. The results file contains the solution of the formulated synthesis problem (such as the selected routes, or materials) as well as the values for all model variables, so they can in addition be used in subsequent stages of the three-stage approach (detailed design and analysis, and innovative and more sustainable design). Figure 6.7 shows the tools integration in the framework.

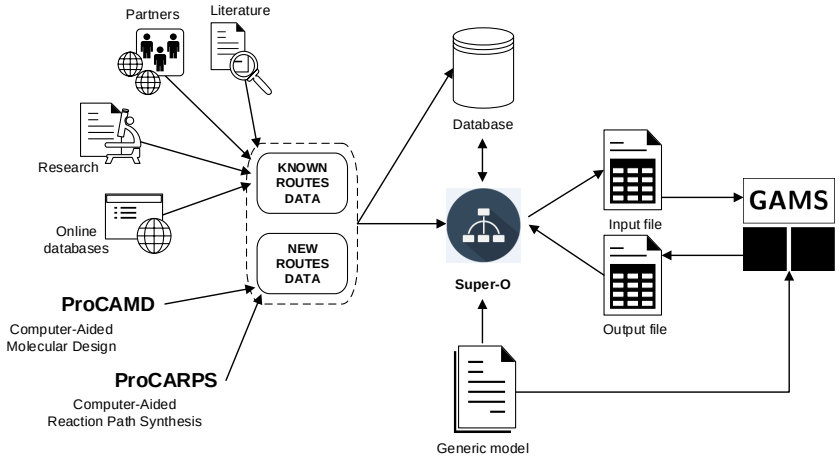


Figure 6.7. Tools integration in the biorefinery synthesis framework: data sources are used to obtain data from well-known or novel routes, which can be stored directly in the database or input in Super-O, from which an input file is automatically generated, the problem is solved in GAMS and an output file is generated and read by Super-O.

In a future vision, the specially designed Biorefinery Synthesis Database described in Chapter 7 and other databases will be integrated in Super-O, enabling the systematic storage and retrieval of the necessary information for different problem types. This way, upon the specification of the problem by the user, all necessary data would be retrieved from databases.

As a result of the systematic structure and the templates generated, standard processing network synthesis problems can be formulated and solved through the user interface of Super-O, without requiring any additional programming. This reduces the time needed for the formulation of this class of problems and, at the same time, it broadens the range of potential users of this optimization-based synthesis methodology, by making it accessible to professionals who are not experts in formulating and solving process synthesis problems using superstructure optimization.

6.3 Conclusion

The workflow and data flow of the framework have been described in detail in this chapter. The workflow consists of 8 steps that guide the user through the problem formulation and the solution of synthesis and/or location-based synthesis problems.

Commercial and in-house tools are integrated with the method, to perform some of the required actions. Moreover, a software implementation of the framework, Super-O, has been presented. Super-O guides users through the solution of synthesis problems by

providing a structured environment for the specification of problem data.

CHAPTER 7

Data management

The data collection step has proven to be one of the most time-consuming steps in the workflow, due to the need to search in different sources and to convert data to a specific format that is compatible with the generic interval model. Therefore, it is crucial that once the data has been collected and transformed, it is stored for its further reuse. To this end, a data structure has been developed, implemented as a database and populated with biorefinery synthesis data.

The objective of the implementation of a data management approach in this work is to provide a common platform for different users to search, store and retrieve data for the formulation and solution of synthesis problems.

This chapter presents an ontology-based data structure, which has been implemented as a Biorefinery Synthesis DataBase (BSDB). Moreover, algorithms for data storage and retrieval are presented. The Biorefinery Synthesis DataBase supports the Biorefinery Synthesis Framework by allowing the systematization, storage, and reuse of synthesis-stage data. The BSDB has been designed, implemented in SQL, and populated with data related to the case studies presented in this thesis as well as other available data.

7.1 Motivation for data management

As highlighted in Chapter 5, the combinatorial nature of the synthesis problem may lead to a large numbers of alternatives with several parameters related to each of them, depending on the models used. Therefore, data management approaches are of crucial importance in order to systematize, store and provide access to these data. This is of special importance in the context of biorefinery synthesis due to the fact that these

processes rely to data coming from various areas of expertise (such as biology/biotechnology, engineering, business, and environmental technology) where different units and standards are common, and location-dependency is an important factor to take into account, hence data for many parameters needs to be collected across locations. Additionally, research on biorefinery processes is currently ongoing and, therefore, standards and databases are not fully established. Finally, another driver towards data management in early-stage decision making is the limited amount of time available to evaluate large numbers of alternatives, which encourages the use of methods and tools, such as databases, to increase efficiency.

A review of available data management approaches and tools in the domain of biorefineries is presented in Section 1.2. Available approaches are valid and hold large amounts of data that has been validated and systematized, however, they are generally not targeted to early-stage synthesis and they generally cover specific sub-domains of data within biorefineries (*e.g.* feedstock composition, thermo-chemical technologies). For these reasons, the integration of these databases with a synthesis method would be cumbersome, as data from multiple databases is required and is often not in the desired format. Therefore, a systematic approach to data management for biorefinery synthesis is necessary.

7.2 Overview of data management

The objective of the implementation of a data management approach in this project is to provide a systematic structure to collect, organize, store, retrieve, update and reuse relevant biorefinery process synthesis data. This task has been addressed in connection to the model blocks presented in Chapter 5, that is, input parameters to the model as well as data required to obtain them are considered as relevant data. The ultimate objective is to create and populate the Biorefinery Synthesis Database for its use in solving biorefinery synthesis problems.

The output of this chapter are a data structure and a database that is specially targeted to early-stage synthesis problems, where data is formatted so that it is compatible with the generic model blocks presented in Chapter 5, yet comprehensive and accessible for other applications. Moreover, the data structure enables automatic data connections to be formed when new items are added, thus providing a platform of continuously growing knowledge. Finally, the data structure is implemented in a biorefinery database, that contains data related to the examples shown in this thesis as well as other biorefinery data. Although a biorefinery implementation is shown in this chapter, the developed data structure is valid for other fields of application, such as wastewater network synthesis, chemical process design, or synthesis of processes for carbon dioxide

utilization.

The approach followed consists of developing an ontology-based data structure (Section 7.3), implementing it as a database (Section 7.4), and developing the necessary algorithms for storing and retrieving data from the developed database (Section 7.4.1).

7.3 Data structure

A data structure has been specifically designed using the method outlined by Singh et al. [10] to support the presented framework in terms of data management. It consists of three main data sections: (i) a general data section (components, locations, reactions), (ii) a section for data related to material (feedstock, products, prices, availability, demand), and (iii) a section containing process data (steps, intervals, tasks, connectivity between intervals). The main parts of the general data section are component, reaction and location data; that is, the component list and pure-component properties, the reaction list and stoichiometry, and list of locations. The material section contains feedstock, intermediate and product data including composition, availability, demand and price. Moreover, it includes utilities and chemicals price data. Finally, the process section includes data on the processing alternatives and connectivity between them. Data on the processing intervals is organized in terms of processing tasks, namely mixing, reaction, waste removal, separation and utility consumption. A scheme of the data structure is presented in Figure 7.1, where the three data sections are represented, along with some of their sub-sections.

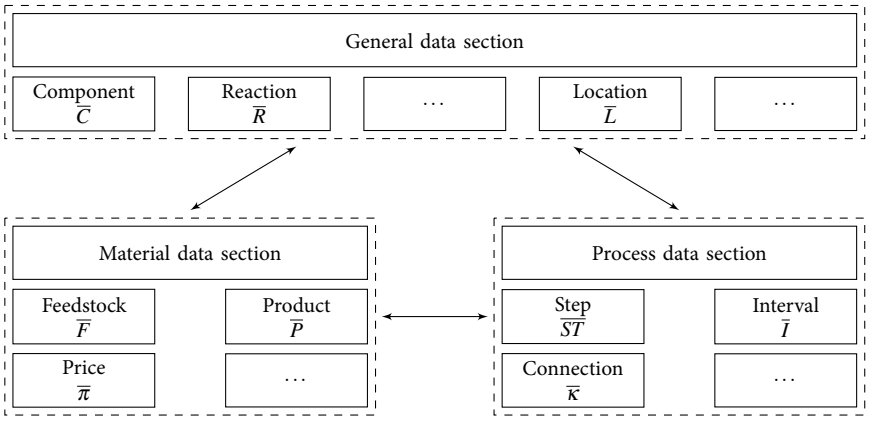


Figure 7.1. The data structure consists of three data sections, general, material and process, which are interconnected.

The three data sections are not independent from each other, instead, there is connections between them. For example, price data in the material section are indexed by

location, using locations from the location list in the basic section, or reactions from the reaction list in the general data section are used in the process data section for reaction task data.

7.3.1 General data section

The general data section contains the following main sub-sections: component, reaction and location. The component sub-section contains component data in terms of name, formula, and pure-component properties (*e.g.*, molar weight). The reaction sub-section includes the reaction names, components involved in them, stoichiometry, and reactions sets into which the reactions are organized. The location sub-section includes a list of locations, for instance countries. Part of the structure of this data section with a selected set of its sub-sections is shown in Figure 7.2. The data classes and instances of this section are listed in Table 7.1.

Note that when retrieving data for the solution of synthesis problems using the generic model, data in the general data sections provide the lists of elements of the main sets: components, reactions, locations. Other sets related to the superstructure are stored in the process section.

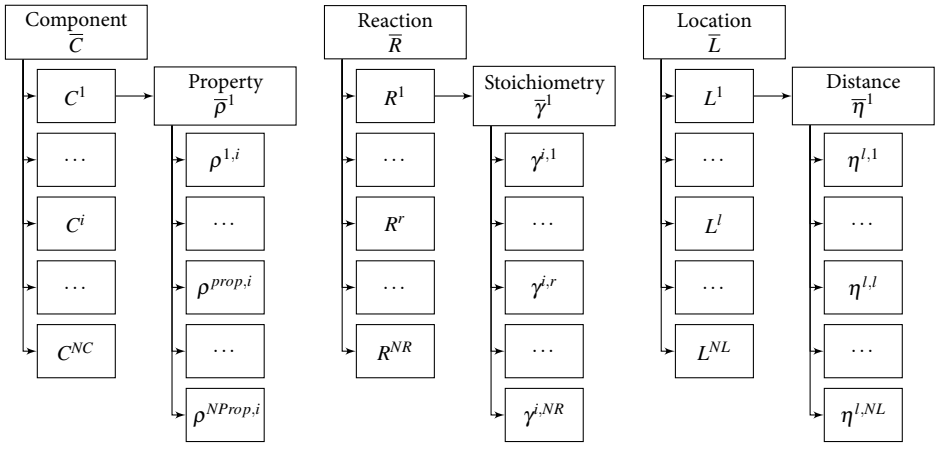


Figure 7.2. Structure of the general data section.

Table 7.1. Main classes of the data structure: general section.

Main classes	Relation with instances	Object descriptions
Component ($\overline{C}$)	$\overline{C} = [C^1 \dots C^i \dots C^{NC}]$	C^i: ith component
Property ($\overline{\rho}$)	$\overline{\rho} = [\overline{\rho}^1 \dots \overline{\rho}^i \dots \overline{\rho}^{NC}]$	$\overline{\rho}^i$: properties of the ith component
	$\overline{\rho^i} = [\rho^{1,i} \dots \rho^{pr,i} \dots \rho^{NPr,NC}]$	$\rho^{pr,i}$: prth property of the ith component
Reaction ($\overline{R}$)	$\overline{R} = [R^1 \dots R^r \dots C^{NR}]$	R^r: rth reaction
Stoichiometry ($\overline{\gamma}$)	$\overline{\gamma} = [\overline{\gamma}^1 \dots \overline{\gamma}^r \dots \overline{\gamma}^{NR}]$	$\overline{\gamma}^j$: stoichiometry of the rth reaction
	$\overline{\gamma^i} = [\gamma^{i,1} \dots \gamma^{i,r} \dots \gamma^{i,NR}]$	$\gamma^{i,r}$: stoichiometric coefficient of ith component in rth reaction
Utility ($\overline{U}$)	$\overline{U} = [U^1 \dots U^{ut} \dots U^{NU}]$	U^{ut}: utth utility
Location ($\overline{L}$)	$\overline{L} = [L^1 \dots L^l \dots L^{NL}]$	L^l: lth location
Distance ($\overline{\eta}$)	$\overline{\eta} = [\overline{\eta}^1 \dots \overline{\eta}^l \dots \overline{\eta}^{NL}]$	$\overline{\eta}^l$: distance from lth location
	$\overline{\eta^l} = [\eta^{l,1} \dots \eta^{l,ll} \dots \eta^{l,NL}]$	$\eta^{l,ll}$: distance from lth location to llth location

7.3.2 Material data section

The material section is sub-divided into the following: feedstock, product, and utility. Data regarding feedstock, their composition, properties, and location-dependent availability and price is stored in the feedstock sub-section. The product sub-section contains products, their specifications, and location-dependent demand and price. Similarly, the utility section contains a list of utilities, their properties and price for different locations. Three sub-sections and their content are shown in Figure 7.3. The main classes and instances of this section are listed in Table 7.2.

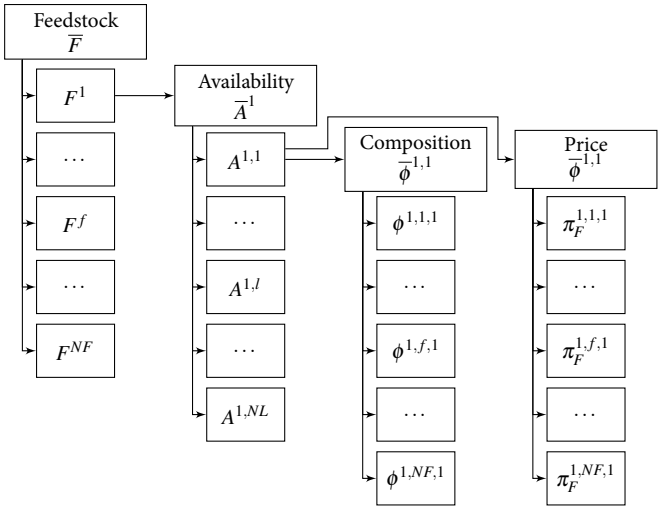


Figure 7.3. Structure of the material data section.

Table 7.2. Main classes of the data structure: material section.

Main classes	Relation with instances	Object descriptions
Feedstock ($\overline{F}$)	$\overline{F} = [F^1 \dots F^f \dots F^{NF}]$	F^f: fth feedstock
Feedstock availability ($\overline{A}$)	$\overline{A} = [\overline{A}^1 \dots \overline{A}^f \dots \overline{A}^{NF}]$	$\overline{A}^f$: availability of the fth feedstock
	$\overline{A}^f = [A^{f,1} \dots A^{f,l} \dots A^{f,NL}]$	$A^{f,l}$: availability of fth feedstock in lth location
Feedstock composition ($\overline{\phi}$)	$\overline{\phi} = [\overline{\phi}^1 \dots \overline{\phi}^f \dots \overline{\phi}^{NF}]$	$\overline{\phi}^f$: composition of the fth feedstock
	$\overline{\phi}^f = [\overline{\phi}^{f,1} \dots \overline{\phi}^{f,l} \dots \overline{\phi}^{f,NL}]$	$\overline{\phi}^{f,l}$: composition of the fth feedstock in the lth location
	$\overline{\phi}^{i,l} = [\phi^{1,f,l} \dots \phi^{i,f,l} \dots \phi^{NC,f,l}]$	$\phi^{i,f,l}$: ith component composition in fth feedstock in the lth location
Feedstock price ($\overline{\pi}_F$)	$\overline{\pi}_F = [\overline{\pi}_F^1 \dots \overline{\pi}_F^f \dots \overline{\pi}_F^{NF}]$	$\overline{\pi}_F^f$: price of the fth feedstock
	$\overline{\pi}_F^f = [\pi_F^{f,1} \dots \pi_F^{f,l} \dots \pi_F{}^{f,NL}]$	$\pi_F^{f,l}$: price of the fth feedstock in the lth location
Product ($\overline{P}$)	$\overline{P} = [P^1 \dots P^p \dots P^{NP}]$	P^p: pth product
Product demand ($\overline{D}$)	$\overline{D} = [\overline{D}^1 \dots \overline{D}^p \dots \overline{D}^{NP}]$	$\overline{D}^p$: demand of the pth product
	$\overline{D}^p = [D^{p,1} \dots D^{p,l} \dots D^{p,NL}]$	$D^{p,l}$: demand of the pth product in the lth location
Product price ($\overline{\pi}_P$)	$\overline{\pi}_P = [\overline{\pi}_P^1 \dots \overline{\pi}_P^p \dots \overline{\pi}_P{}^{NP}]$	$\overline{\pi}_P^p$: price of the pth product
	$\overline{\pi}_P^p = [\pi_F^{p,1} \dots \pi_F^{p,l} \dots \pi_F^{p,NL}]$	$\pi_F^{p,l}$: price of the pth product in the lth location
Chemical price ($\overline{\pi}_C$)	$\overline{\pi}_C = [\overline{\pi}_C^1 \dots \overline{\pi}_C^i \dots \overline{\pi}_C^{NC}]$	$\overline{\pi}_C^i$: price of the ith chemical
	$\overline{\pi}_C^i = [\pi_C^{i,1} \dots \pi_C^{i,l} \dots \pi_C^{i,NL}]$	$\pi_C^{i,l}$: price of the ith chemical in the lth location
Utility price ($\overline{\pi}_U$)	$\overline{\pi}_U = [\overline{\pi}_U^1 \dots \overline{\pi}_U^{ut} \dots \overline{\pi}_U^{NU}]$	$\overline{\pi}_U^{ut}$: price of the utth utility
	$\overline{\pi}_U^{ut} = [\pi_U^{ut,1} \dots \pi_U^{ut,l} \dots \pi_U^{ut,NL}]$	$\pi_U^{ut,l}$: price of the utth utility in the lth location

7.3.3 Process data section

The process section contains data for the following: process step, interval, connection, and tasks within each step (mixing, reaction, waste removal, product separation, and utility consumption). The process steps sub-section has a list of processing steps and their relative position (*i.e.*, before/after other steps). Alternatives within a process step are represented by different processing intervals, listed in the interval sub-section. The connections sub-section contains intervals connectivity data, which is given through inlet-outlet specifications. Therefore, superstructure connections for a specific problem can be inferred by comparison of inlets and outlets of intervals. Each task sub-section contains specific data for the given task, for example reaction conversion or separation recovery factor. The main structure of this section is shown in Figure 7.4. The different data levels in the structure are depicted in Figure 7.5. The main classes of this section are listed in Tables 7.3 and 7.4.

Table 7.3. Main classes of the data structure: process section.

Main classes	Relation with instances	Object descriptions
Processing step ($\overline{ST}$)	$\overline{ST} = [ST^1 \dots ST^{st} \dots ST^{NS}]$	ST^{st}: stth processing step
Processing interval ($\overline{I}$)	$\overline{I} = [I^1 \dots I^k \dots I^{NI}]$	I^k: kth processing interval
Inlet material (κ_{IN})	$\overline{\kappa}_{IN} = [\overline{\kappa}^1_{IN} \dots \overline{\kappa}^k_{IN} \dots \overline{\kappa}_{IN}]$	$\overline{\kappa}^k_{IN}$: inlet material to the kth interval
	$\overline{\kappa}^k_{IN} = [\kappa^{k,1} \dots \kappa^{k,in}_{IN} \dots \kappa^{k,NIn}_{IN}]$	$\kappa^{k,in}_{IN}$: inth inlet to the kth interval
Outlet material (κ_{OUT})	$\overline{\kappa}_{OUT} = [\overline{\kappa}^1_{OUT} \dots \overline{\kappa}^k_{OUT} \dots \overline{\kappa}_{OUT}]$	$\overline{\kappa}^k_{OUT}$: outlet material to the kth interval
	$\overline{\kappa}^k_{OUT} = [\kappa^{k,1} \dots \kappa^{k,out}_{OUT} \dots \kappa^{k,NOut}_{OUT}]$	$\kappa^{k,out}_{OUT}$: outth outlet to the kth interval
Connection ($\overline{\zeta}$)	$\overline{\zeta} = [\overline{\zeta}^1 \dots \overline{\zeta}^k \dots \overline{\zeta}^{NI}]$	$\overline{\zeta}^k$: connection from the kth interval
	$\overline{\zeta_k} = [\zeta^{k,1} \dots \zeta^{k,kk} \dots \zeta^{k,NI}]$	$\zeta^{k,kk}$: connection from the kth interval to the kkth interval
Primary connection ($\overline{\zeta_P}$)	$\overline{\zeta}_P = [\overline{\zeta}^1_P \dots \overline{\zeta}^k_P \dots \overline{\zeta}^{NI}_P]$	$\overline{\zeta}^k_P$: primary connection from the kth interval
	$\overline{\zeta}^k_P = [\zeta^{k,1}_P \dots \zeta^{k,kk}_P \dots \zeta^{k,NI}_P]$	$\zeta^{k,kk}_P$: primary connection from the kth interval to the kkth interval

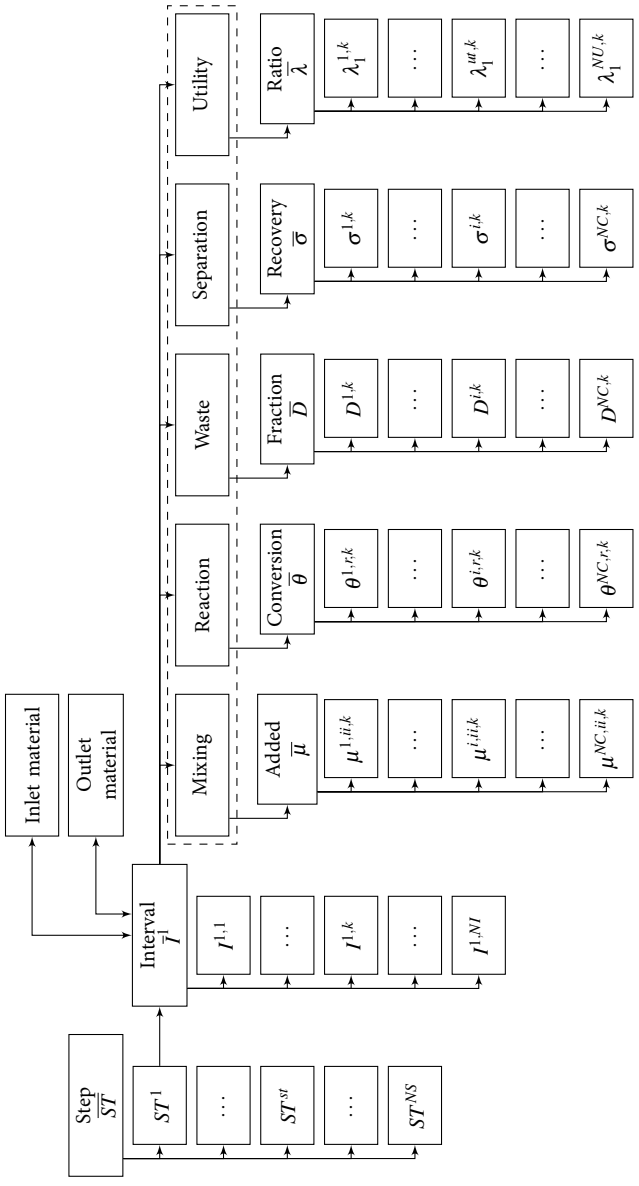


Figure 7.4. Structure of the process data section.

Table 7.4. Sub-classes in the process data section: processing tasks.

Main classes	Relation with instances	Object descriptions
Mixing task		
Added chemical ratio ($\overline{\mu}$)	$\overline{\mu} = [\overline{\mu}^1 \dots \overline{\mu}^k \dots \overline{\mu}^{NI}]$	$\overline{\mu}^k$: ratio of added chemicals to the kth interval
	$\overline{\mu}^k = [\overline{\mu}^{1,k} \dots \overline{\mu}^{i,k} \dots \overline{\mu}^{NC,k}]$	$\overline{\mu}^{i,k}$: ratio of ith added chemical to the kth interval
	$\overline{\mu}^{i,k} = [\mu^{i,1,k} \dots \mu^{i,ii,k} \dots \mu^{i,NC,k}]$	$\mu^{i,ii,k}$: ratio of ith chemical added to the kth interval with respect to the iith chemical inlet flow rate
Reaction task		
Conversion ($\overline{\theta}$)	$\overline{\theta} = [\overline{\theta}^i \dots \overline{\theta}^k \dots \overline{\theta}^{NI}]$	$\overline{\theta}^k$: reaction conversion in the kth interval
Waste removal task		
Waste fraction ($\overline{\delta}$)	$\overline{\delta} = [\overline{\delta}^1 \dots \overline{\delta}^k \dots \overline{\delta}^{NI}]$	$\overline{\delta}^k$: waste fraction in the kth interval
	$\overline{\delta}^k = [\delta^{1,k} \dots \delta^{i,k} \dots \delta^{NC,k}]$	$\delta^{i,k}$: waste fraction of the ith component in the kth interval
Separation task		
Recovery ($\overline{\sigma}$)	$\overline{\sigma} = [\overline{\sigma}^1 \dots \overline{\sigma}^k \dots \overline{\sigma}^{NI}]$	$\overline{\sigma}^k$: recovery in the separation task of the kth interval
	$\overline{\sigma}^k = [\sigma^{1,k} \dots \sigma^{i,k} \dots \sigma^{NC,k}]$	recovery of ith component in the kth interval (first consumption point)
Utility consumption task		
Utility ratio ($\overline{\lambda}_1$)	$\overline{\lambda}_1 = [\overline{\lambda}_1^1 \dots \overline{\lambda}_1^k \dots \overline{\lambda}_1^{NI}]$	ratio of utility consumed in the kth interval (first consumption point)
	$\overline{\lambda}_1^k = [\lambda_1^{1,k} \dots \lambda_1^{ut,k} \dots \lambda_1^{NU,k}]$	ratio of utth utility consumed in the kth interval (first consumption point)

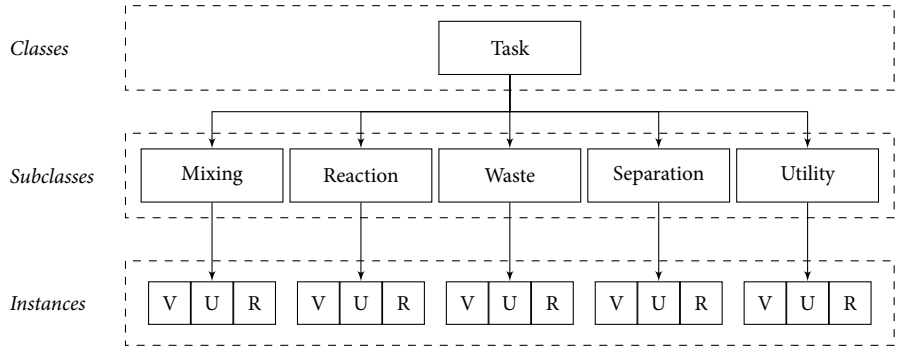


Figure 7.5. Representation of the systematization of data for tasks in each interval in terms of classes, subclasses and instances. Nomenclature: V is value (numerical), U is unit of measure, and R is reference.

7.3.4 Data relations

Data tables within each section as well as across sections are related to each other through data relations, the implementation of which allows data consistency and facilitates data search and retrieval. For example, a reaction component needs to be in the component list, an interval for which a connection is given should be in the interval list, or in order to define the price for a raw material in a geographic location, it needs to have an availability larger than zero in that specific location.

Connections between intervals in the data structure are data connections automatically defined as data is input. Each interval is defined with a set of input and output materials, that are used to create connections. When an interval has an output material that acts as input material to another interval, a connection is created between them, so that when a superstructure with these intervals is retrieved, a superstructure connection exists between them. Note that when a separation exists in an interval, outlet materials are be defined as primary or secondary outlets, thus giving rise to primary or secondary connections.

7.4 Biorefinery Synthesis Database

The Biorefinery Synthesis Database (BSDB) has been developed with the objective to allow the storage and reuse of relevant biorefinery synthesis data. The database has been build upon the data structure presented in Section 7.3. This section presents the implementation and content of said database.

An SQL implementation of the biorefinery database has been created based on the presented data structure (see Section 7.3). Data tables are created for classes and subclasses and data relations are specified to ensure consistency. The overall structure of the database is shown in Figure 7.6, where data examples are shown, illustrating relations across data sections.

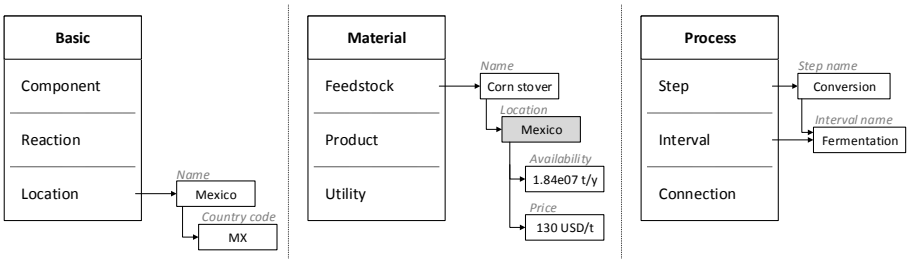


Figure 7.6. Data structure implemented as the Biorefinery Synthesis Database (BSDB) developed to support the biorefinery synthesis framework including data examples.

Statistics related to the data available in the Biorefinery Synthesis Database are listed in Table 7.5. Data used to populate the BSDB corresponds to data from examples solved using the biorefinery synthesis framework, a non-exhaustive overview of some of the routes available in the database is given in Section 7.4.1 on page 102.

Table 7.5. Biorefinery Synthesis Database statistics.

Data	Biorefinery Database
Components	71
Utilities	5
Processing steps	21
Processing intervals	102
Feedstocks	11
Products	9
Reactions	63
Locations	10

7.4.1 Search engines

Search engines have been developed to automate the retrieval of synthesis data from the Biorefinery Synthesis Database. A variety of queries can be used to retrieve data from the database. For instance, a single technology and its performance, the whole set of alternatives for a processing step, pricing for a given biomass-based feedstock, location-availability of biomass, market price of products, etc.

One of the relevant searches in the database is the search for processing superstructures. That is, given a (set of) feedstock(s) and/or product(s) and other search constraints, a superstructure of alternatives is to be retrieved. In the BSDB, these superstructures are generated from intervals in the database by finding connections between processing intervals based on their predefined inlet and outlet materials. Networks may be generated starting from a feedstock or a product, the former case is referred to as forward network generation (see Figure 7.7), whereas the latter is called backward network generation (see Figure 7.8).

Algorithm for forward network generation

The algorithm for forward network generation shown in Figure 7.7 takes a feedstock identifier and looks for the feedstock interval to finds its outlet identifier, which is used to find intervals that take it as inlet. If such intervals are found, they are stored and their oulets are taken, to search for intervals that take them as inputs. This procedure continues until one or more product intervals are reached. The output is a superstructure of processing intervals that take the pre-defined feedstock as inlet and convert it to a set of products through a sequence of processing steps with alternative processing intervals for each of them.

Algorithm for backward network generation

Similarly to the forward algorithm, the backward search algorithm shown in Figure 7.7 generates superstructure networks from the Biorefinery Synthesis Database taking product identifiers as input. The output is, similarly to the forward algorithm, a superstructure of alternatives leading to the desired product. The superstructure is expected to include all relevant feedstocks that can lead to the given product, except if criteria have been specified as search constraints (for example, exclude lignocellulosic feedstocks, or include only feedstocks available a Brazil). Moreover, besides the pre-defined product, the superstructure might include any co-products being produced along the processing network. For instance, if butanol is specified as product, ethanol and acetone are produced in some routes as co-products, and they should appear in the output superstructure.

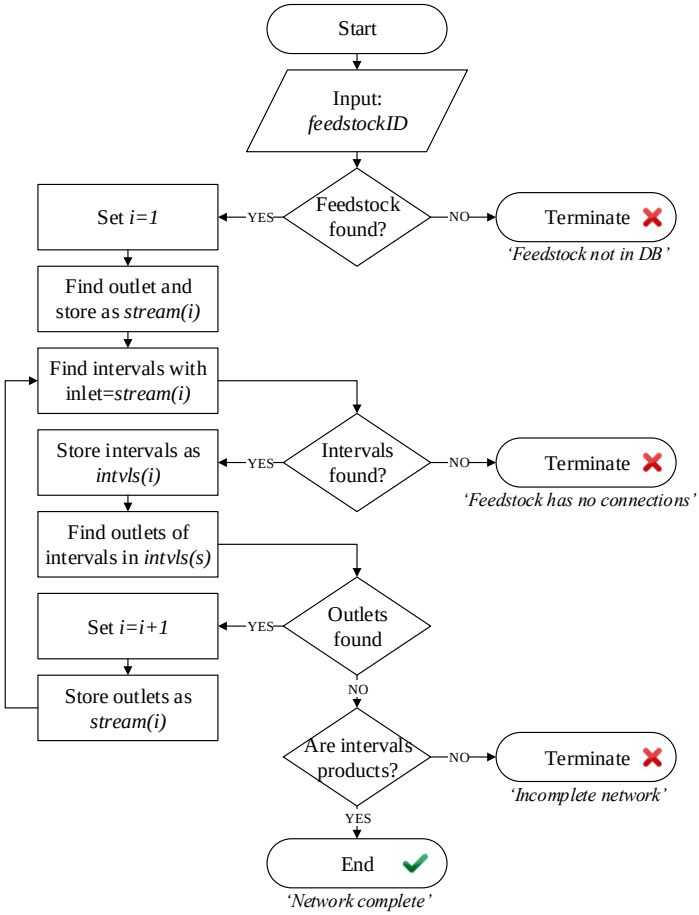


Figure 7.7. Forward network generation algorithm.

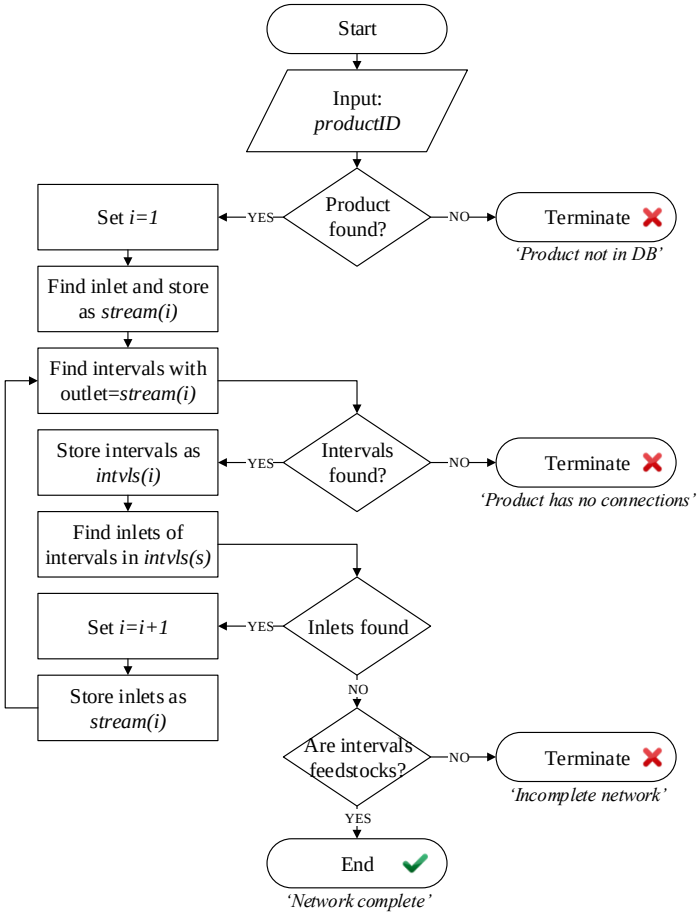
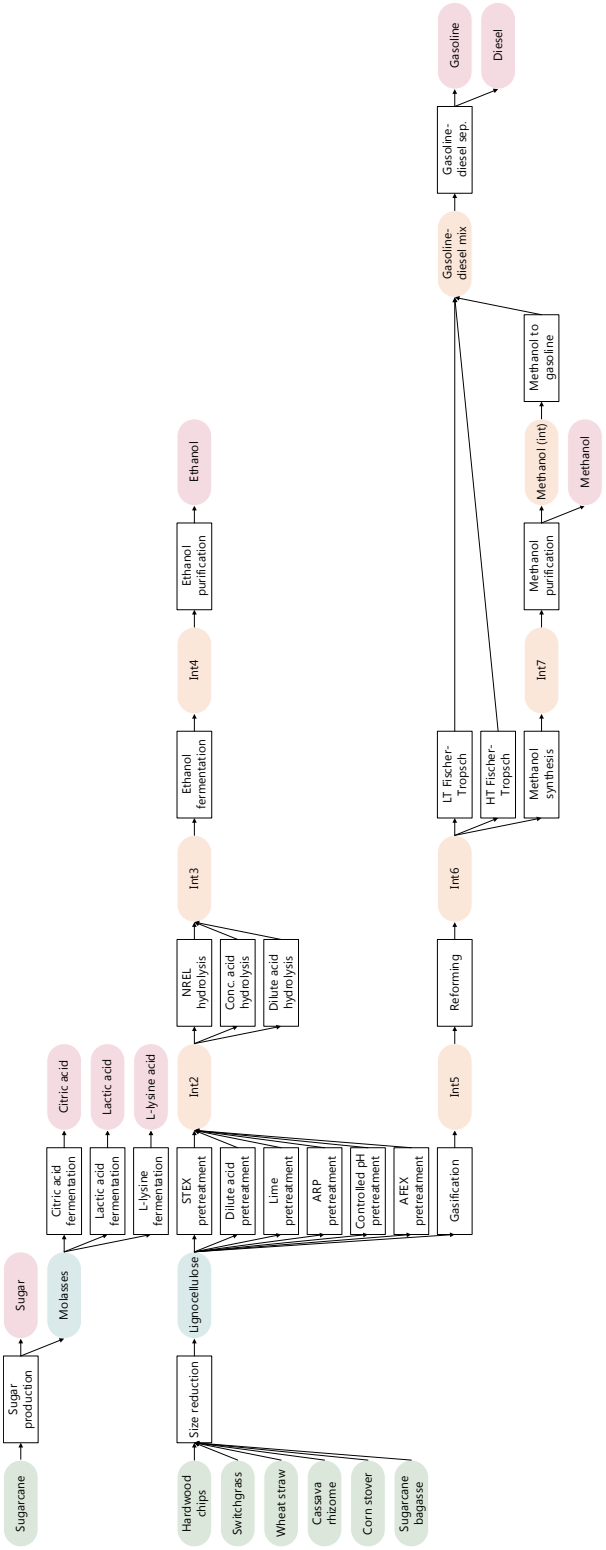


Figure 7.8. Backward network generation algorithm.


Sugarcane
Sugar production
Sugar
Molasses
Citric acid fermentation
Lactic acid fermentation
L-lysine fermentation
Citric acid
Lactic acid
L-lysine acid
Hardwood chips
Switchgrass
Wheat straw
Cassava rhizome
Corn stover
Sugarcane bagasse
Size reduction
Lignocellulose
STEX pretreatment
Dilute acid pretreatment
Lime pretreatment
ARP pretreatment
Controlled pH pretreatment
AFEX pretreatment
Int2
NREL hydrolysis
Conc. acid hydrolysis
Dilute acid hydrolysis
Int3
Ethanol fermentation
Int4
Ethanol purification
Ethanol
Gasification
Int5
Reforming
Int6
LT Fischer-Tropsch
HT Fischer-Tropsch
Methanol synthesis
Int7
Methanol purification
Methanol (int)
Methanol to gasoline
Methanol
Gasoline-diesel mix
Gasoline-diesel sep.
Gasoline
Diesel

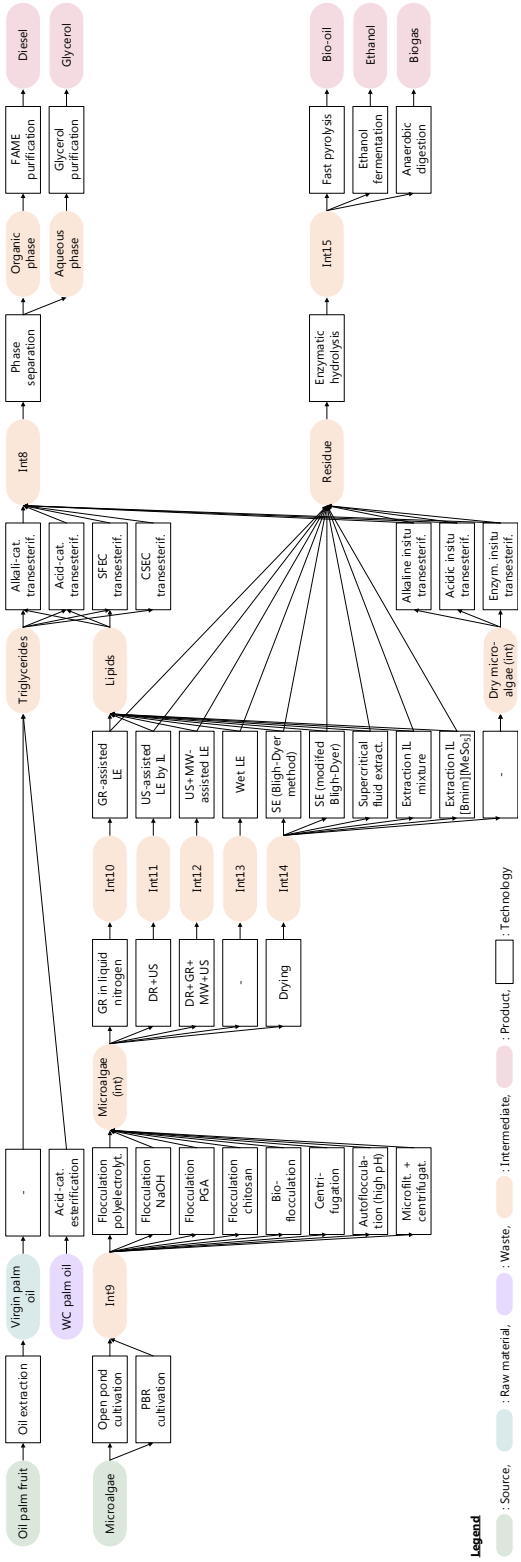


Figure 7.8. Pathways in the Biorefinery Synthesis Database.

7.5 Conclusion

A data structure has been developed so as to assist the collection, systematization, storage, retrieval, update and reuse of synthesis data. This data structure can be applied to a wide range of applications. The synthesis of biorefineries has been considered in this thesis, thus an implementation of the data structure as the Biorefinery Synthesis Database has been described. Moreover, search engines for the retrieval of superstructure data have been presented.

Although only deterministic solutions are considered in this thesis, stochastic solution approaches have been developed and may be applied to this type of problems [87]. For this reason, uncertainty data, if available, has been stored in the database, in terms of data ranges or standard deviation.

Moreover, market data, which experiences variability over time, has been updated to a reference year (2015) and historical data is stored, if available, for its use in forecasting.

PART III

Application examples

Part III of this thesis illustrates the application of the developed framework to examples within the area of biorefinery networks synthesis. Three examples are presented in terms of problem type and in order of increasing complexity: the first example is a *process-product network synthesis* problem, the second example is a *feedstock-process-product network synthesis* problem, and the third and last example is a *location-dependent feedstock-process network synthesis* problem.

The development of the framework and its application to relevant examples were carried out simultaneously. The obtained results and identified limitations from the examples were used to drive further developments of the framework. For this reason, the entire framework in its full potential is not applied to all examples. The first example was used to understand the abilities of the initial steps of the framework (with no location-dependency). The second example has an increased complexity and starts exploring location dependencies. The third example is first used to understand the location dependency of biorefinery synthesis solutions and the fully extended framework is then applied to cover location-dependent synthesis including transportation.

In terms of application, the first example concerns the production of value-added chemicals from sugarcane molasses. The production of biodiesel and co-products are the focus of the second example, where two sets of feedstocks are considered: vegetable oil and microalgae. The third case study targets the production of bioethanol from various lignocellulosic feedstocks.

Various scenarios are considered in each example. For clarity, a consistent nomenclature for scenarios is used throughout the thesis. Scenarios related to the first example, utilization of molasses, are represented starting with “M”, those related to the production of diesel with “D”, and “E” is used for scenarios related to ethanol production, the third example.

CHAPTER 8

Synthesis of product and process networks

– Sugarcane molasses utilization

The synthesis of process and product networks corresponds to problems where feedstock(s) are fixed and an array of products with their corresponding routes are considered. This problem type is relevant when a feedstock is available, for example, as a by-product or residue of another process. The structure of this problem type is shown in Figure 8.1.

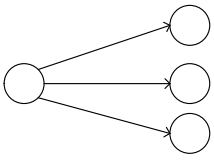


Figure 8.1. Process and product network synthesis problem.

This example involves the utilization of sugarcane molasses, a by-product of the sugar industry, in Mexico. This problem is considered of interest due to its relevance given the molasses surplus in this location. Moreover, it serves as a simple example to test the application of the basic framework (without location-dependency) and identify needs for its further development. This problem was initially investigated by Anaya-Reza et al. [88] from a detailed design perspective. A collaboration was the results of this work, in which data was shared to test the application of the synthesis framework.

8.1 Motivation for the utilization of sugarcane molasses

Sugarcane molasses are a by-product of sugar production from sugarcane, from which no more sugar can be obtained via physical methods. Molasses have various end uses, which include animal feed, alcohol production, industrial re-processing to extract the remaining sugar, and industrial production [88].

The bioconversion of sugarcane molasses to products for the food industry is the focus of this case study located in Mexico. This work is the result of a collaboration with UAM [88].

8.2 Formulation and solution

The formulation and solution of a synthesis problem for the production of value-added chemicals from sugarcane molasses using the presented framework is shown in this section. This is a simple problem which is used to test the framework.

8.2.1 Step 1: Problem definition

This problem deals with the utilization of a stream of 40 kt/y of sugarcane molasses, a by-product of sugarcane processing, and its conversion to chemicals for the food industry.

Step 1.1: Objective and scope definition

Action 1.1.1: The objective is to determine the optimal product derived from sugarcane molasses so that the economic profit is maximized. The effect of using different objective functions is considered in this example.

Action 1.1.2: The following are defined:

a) Feedstock: sugarcane molasses
b) Technogies: biochemical conversion
c) Location: Mexico

Step 1.2: Solution scenarios description

Action 1.2.1: A single location (Mexico) is considered for this example and supply chain and transportation considerations are not included. Therefore, only the basic synthesis problem is to be solved, without including location selection.

Action 1.2.2: Three scenarios are taken into account in terms of the considered objective function:

- Scenario M-1: operating profit maximization (GOI without utilities)
- Scenario M-2: operating profit maximization (GOI with utilities)
- Scenario M-3: EBIT maximization (including capital costs)

It should be noted that the three considered objective functions are variations of the EBIT profit function. Therefore, the three scenarios can be solved by actually modifying the objective function in the model, or simply by setting certain parameters to zero, when a given term in the objective function is not to be considered. For example, for scenarios without capital costs (scenarios M-1 and M-2), capital cost parameters can be set to zero, which makes the capital cost term take a value of zero, hence reducing the EBIT function to operating profit.

Action 1.2.3: Single optimal solutions are considered in this case given the relatively small number of alternative solutions, hence integer cuts are not included in the model file.

8.2.2 Step 2: Alternatives definition

A set of bioproducts derived from sugarcane molasses have been pre-selected based on their use in the food industry and their national demand in Mexico [88]. These products are: L-lysine, lactic acid, and citric acid.

Step 2.1: Superstructure generation

Action 2.1.1: A database search using the forward network generation with the feedstock (sugarcane molasses) as input did not return results, since this was the first case study performed in this project.

Action 2.1.2: A superstructure of alternatives is generated based on alternatives data collected from Anaya-Reza et al. [88]. The superstructure is shown in Figure 8.2, which consists of a sequence of processing steps listed in Table 8.1.

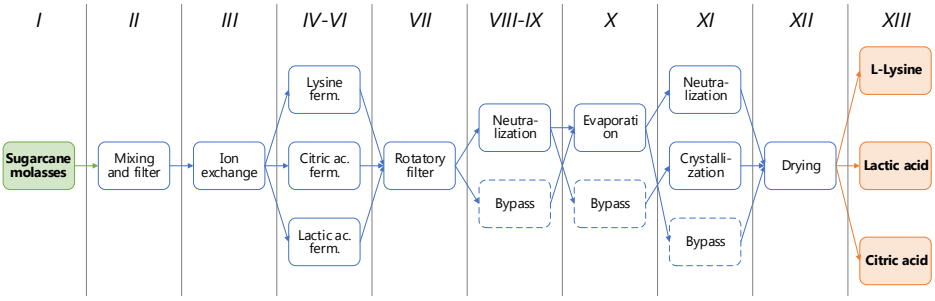


Figure 8.2. Superstructure of alternatives for the sugarcane molasses example.

Table 8.1. List of processing steps in the superstructure for molasses utilization.

Step	Description
I	Raw material
II	Dilution and solids removal
III	Impurities removal
IV-VI	Fermentation
VII	Biomass separation
VIII	Neutralization 1
IX	Neutralization 2
X	Concentration
XI	Crystallization
XII	Drying
XIII	Product

Step 2.2: Data collection

Action 2.2.1: Alternatives from the database are not used.

Action 2.2.2: Data for all alternatives is collected. Based on the collected data, the superstructure is redefined to incorporate connectivity rules between intervals and to ensure the proper representation of alternatives. For example, since reactions in series are present in the fermentation intervals, several processing intervals are used in series to represent a single fermentation step. Moreover, some alternatives considered perform differently depending on the input stream conditions, hence multiple intervals are used to capture this. The superstructure after this step is depicted in Figure 8.3 and processing intervals are listed Table 8.2.

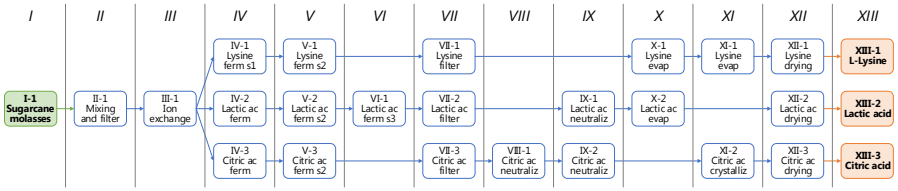


Figure 8.3. Superstructure of alternatives for the sugarcane molasses example after Action 2.2.2. Reactions are divided into three steps to account for reactions in series and technologies that exhibit different performance in different routes are represented by multiple intervals.

Feed composition data, reactions and feedstock/product price data are given in Tables 8.3 to 8.5. A more detailed representation of the superstructure for this problem is depicted in Figure B.1 on page 190.

Table 8.2. Processing intervals for the production of value-added products from sugarcane molasses.

Interval	Description
I-1	Sugarcane molasses
II-1	Mixing and filter
III-1	Ion exchange
IV-1	L-lysine fermentation step 1
IV-2	Lactic acid fermentation step 1
IV-3	Citric acid fermentation step 1
V-1	L-lysine fermentation step 2
V-2	Lactic acid fermentation step 2
V-3	Citric acid fermentation step 2
VI-1	Lactic acid fermentation step 3
VII-1	L-lysine filter
VII-2	Lactic acid filter
VII-3	Citric acid filter
VIII-1	Citric acid neutralization
IX-1	Lactic acid neutralization
I-1	Sugarcane molasses

Table 8.3. Sugarcane molasses composition [88].

Component	Composition [mass fraction]	Simplified composition [mass fraction]
ash	0.12	–
glucose	0.07	0.15
impurities	0.004	0.01
levulose[a]	0.09	–
nitrogenated compounds	0.095	–
non-fermentable sugars	–	0.20
other carbohydrates	0.041	–
reducing sugars	0.03	–
sucrose[b]	0.35	0.35
water	0.20	0.29

[a] fructose [b] saccharose

Table 8.4. Reaction stoichiometry data for the molasses example. Rows are components and columns are reactions, where L-reactions are related to the conversion of molasses to L-lysine, CA-reactions to citric acid, and LA-reactions to lactic acid.

	L-1	L-2	L-3	CA-1	CA-2	CA-3	CA-4	CA-5	LA-1	LA-2	LA-3	LA-4
sucrose	-1											
water	-1											
glucose	2											
glucose		-1										
ammonia		-1.12										
oxygen		-4.08										
soy-hydrolysate		-0.08										
biomass		1.24										
carbon-dioxide		3.16										
lysine		0.6										
organic-acids		0.48										
water		4.35										
lysine			-1									
HCl			-1									
lysine-HCl			1									
sucrose				-1								
water				-1								
glucose				2								
glucose					-1							
amm-sulfate					-0.02							
nutrients					-0.06							
oxygen					-1.75							
biomass					0.65							
carbon-dioxide					0.68							
citric-acid					0.8							
water					2.78							
citric-acid						-1						
ca-hydroxide						-1.5						
ca-citrate						0.5						
water						3						
ca-citrate							-1					
sulfuric-acid							-3					
gypsum							3					
citric-acid							2					
citric-acid								-1				
ca-crystal								1				
sucrose									-1			
water									-1			
glucose									2			
glucose										-1		
nutrients										-0.06		
biomass										0.49		
lactic-acid										2		
ca-hydroxide											-1	
lactic-acid											-2	
ca-lactate											1	
water											1	
ca-lactate												-1
sulfuric-acid												-1
gypsum												1
lactic-acid												2

Table 8.5. Price data for feedstock and product intervals [88].

Interval	Description	Price	Unit
I-1	Molasses	150	$/t
XIII-1	L-lysine	2600	$/t
XIII-2	Lactic acid	2000	$/t
XIII-3	Citric acid	800	$/t

Action 2.2.4: Data is input in Super-O.

Action 2.2.3: Collected data is stored in the database for future reference and reuse.

8.2.3 Step 3: Modeling

Step 3.1: Model setup

Action 3.1.1: The objective function is EBIT, as defined in Equation (5.53). It should be noted that capital costs include only equipment purchase costs.

Action 3.1.2: Model blocks corresponding to the basic model formulation are selected for this example.

Step 3.2: Definition of scenarios and model inputs

Action 3.2.1: Scenario-dependent model inputs are listed in Table 8.6.

Table 8.6. Values of scalar model inputs for sugarcane example.

Scalar	Value	Unit
FeedstockRate	40	kt/y
BigM	1.00E+06	–
WasteHandlingPrice	0	$/t
ProjectLifetime	10	y

Action 3.2.2: Input files are generated for Scenarios M-1, M-2, M-3.

8.2.4 Step 4: Single-location solution

Step 4.1: Input file generation and problem solution

Action 4.1.1: Solution algorithm is selected as CPLEX, given the linearity of the problem.

8.3 Results

Results for each scenario of this example are given in Tables 8.7 and 8.8 and Figure 8.4. Table 8.7 provide information on the selected process topology for each scenario, whereas Table 8.8 and Figure 8.4 list the objective function value (EBIT) for each scenario, as well as a breakdown of revenue and cost items. Additional results and data are provided in Appendix B.

Table 8.7. Selected routes for each scenario in sugarcane molasses example.

		Scenario M-1	Scenario M-2	Scenario M-3
Feedstock	**I**	I-1	I-1	I-1
Route	**II**	II-1	II-1	II-1
	III	III-1	III-1	III-1
	IV	IV-2	IV-2	IV-1
	V	V-2	V-2	V-1
	VI	VI-2	VI-2	–
	VII	VII-2	VII-2	VII-1
	VIII	–	–	–
	IX	IX-1	IX-1	–
	X	X-2	X-2	X-1
	XI	–	–	XI-1
	XII	XII-2	XII-2	XII-1
Product	**XIII**	XIII-2	XIII-2	XIII-1
Name		Lactic acid	Lactic acid	L-lysine

Table 8.8. Objective function breakdown for each scenario in sugarcane molasses example.

Term	**Scenario M-1**		**Scenario M-2**		**Scenario M-3**	
Product sales	31.92	M$/y	31.92	M$/y	30.35	M$/y
Raw material cost	6.00	M$/y	6.00	M$/y	6.00	M$/y
Added chemicals cost	4.10	M$/y	4.10	M$/y	4.70	M$/y
Utilities cost	–	–	2.71	M$/y	0.63	M$/y
Capital cost	–	–	–	–	3.16	M$
Profit	21.82	M$/y	19.11	M$/y	18.70	M$/t

The selected product for scenarios M-1 and M-2 is lactic acid whereas adding capital costs to the objective function changes the selected product to L-lysine. This shows that if the problem was a retrofit case, where equipment is available and is to be re-purposed, then the most profitable product operating profit wise would be lactic acid whereas the conclusion changes when considering a greenfield project. It should be noted that the

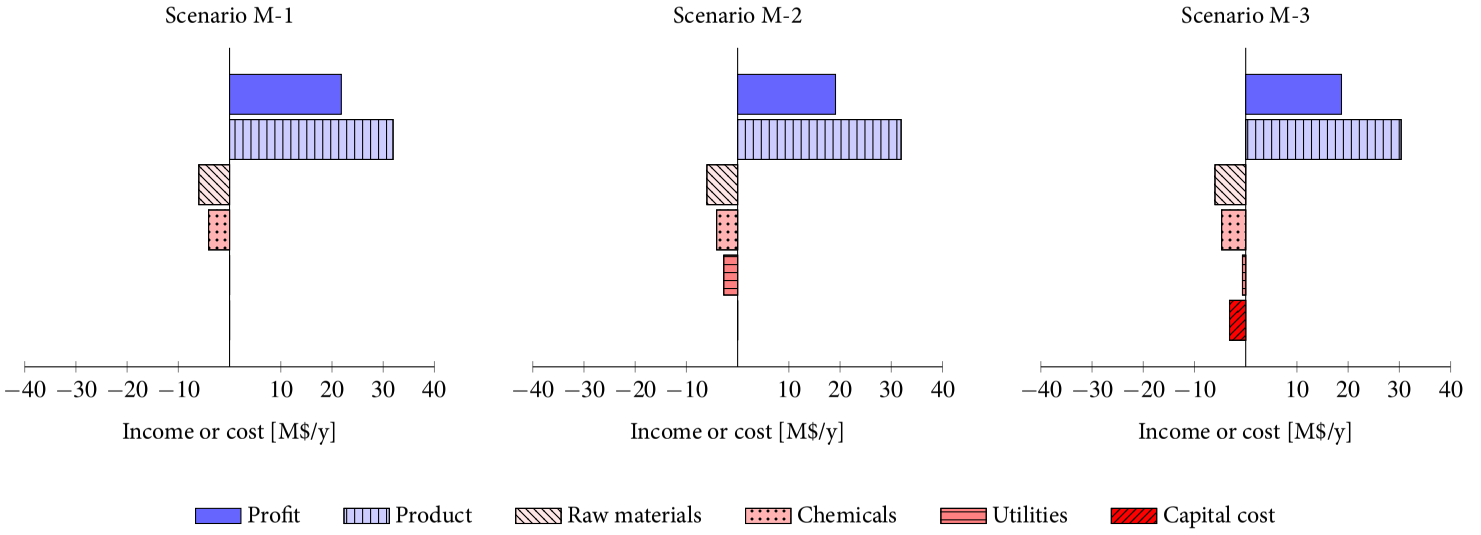


Figure 8.4. Profit, revenue and cost breakdown corresponding to the solution of scenarios M-1, M-2, and M-3. Note that positive values indicate positive profit and revenue, and negative values indicate cost. Positive profit and revenue are further indicated by blue shaded bars, whereas costs can be identified as red shaded bars. Capital cost is shown on a per-year basis, distributed over a 10 year project lifetime.

relatively low value of the capital cost term is due to the low throughput in the process as well as to the fact that merely equipment costs are included in it.

Once the desired product is selected, detailed simulation and analyses should be perform to validate and extend the design of the selected processing route. Techno-economic analyses of these processes have been performed by Anaya-Reza et al. [88].

8.4 Conclusion

The production of value-added products from sugarcane molasses, a by-product of the sugar industry, has been addressed in this example. This simple example is used to demonstrate the applicability of the Biorefinery synthesis framework through its user interface Super-O. Moreover, the influence of choice of objective function in the selected topology has been considered.

The solutions obtained are valid only for the pre-defined geographical location, given that some of the problem data is location-dependent and that the market context determines the alternatives in terms of feedstocks and products. The framework has since been extended in order to include these and other relevant considerations, which are shown in Chapters 9 and 10.

CHAPTER 9

Synthesis of feedstock, process and product networks

– Biodiesel production

The problem of feedstock, process and product network synthesis is highly relevant to the biorefinery area, since biorefineries are integrated facilities where a portfolio of products is produced and selecting from relevant available feedstocks is necessary. The graphical representation of this problem type is shown in Figure 9.1.

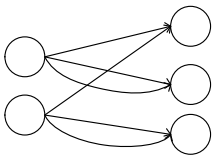


Figure 9.1. Multi-product, multi-product synthesis problem.

The production of biodiesel and co-products is the focus of this example, which leads to a more complex problem with recycle streams, and data from multiple locations.

9.1 Motivation for the production of biodiesel

Biodiesel is an alternative fuel which consists of monoalkyl esters of long chain fatty acids derived from renewable lipid feedstock [89]. The most common way to produce it is by transesterification of lipids, which can be catalyzed (by alkali, acid or enzymatic catalyst) or non-catalyzed. The aforementioned chemical reaction involves a molecule of triacylglycerol (triglyceride from the oil source) reacting with three molecules of an alcohol (for example methanol) to yield three a fatty acid alkyl ester (FAME) and one molecule of glycerol [89].

The production of biodiesel and co-products from different feedstocks is considered in this chapter to illustrate the use of the methodology.

Note that synthesis-design data from previous work in the group on biodiesel production was used to populate the BSDB [90] previously, which is retrieved and expanded in this example.

9.2 Formulation and solution

The formulation and solution of a biodiesel and co-products production synthesis problem from various feedstocks through the application of the step-by-step workflow presented in Chapter 6 is described in this section.

9.2.1 Step 1: Problem definition

In the problem definition step, objectives, scope and scenarios are set for the problem. The following questions are of concern in this example:

Q1. Which feedstock out of the considered set should be used to produce the desired product?
Q2. Which is the optimal processing route (*i.e.*, the route that maximizes/minimizes the predefined objective function)?
Q3. How does including capital costs affect the solution?

Step 1.1: Objective and scope definition

Action 1.1.1: The objective of this problem is to determine the optimal raw material and topology of a processing network for biodiesel production by maximizing the operating profit calculated from sales of product, purchase of raw materials and consumption of chemicals, solvents and utilities.

Action 1.1.2: The following are defined for this problem:

a) Product: biodiesel (and co-products)
b) Locations: Malaysia and South Korea

Action 1.1.3: The following process performance indicators are calculated: kg-raw material/kg-product, kg-waste/kg-product, kg-water/kg-product, op-cost/kg-product.

Step 1.2: Solution scenarios definition

Action 1.2.1: Location and transportation issues are not directly included in the solution. This means that location selection is not performed via optimization though

location selection variables. However, some data is location-dependent, such as prices, for which Malaysia and South Korea are considered as locations. That is, although the selection of locations is not directly performed, if microalgae are selected as feedstock, since their data is related to South Korea, this becomes the biorefinery location.

Action 1.2.2: Two scenarios are defined:

- Scenario D-1: Synthesis of feedstock-process-product network so that GOI is maximized (without considering utilities)
- Scenario D-2: Synthesis of feedstock-process-product network so that GOI is maximized (including utilities)
- Scenario D-3: Synthesis of feedstock-process-product network so that EBIT is maximized

Action 1.2.3: Single solutions are considered for each scenario.

9.2.2 Step 2: Alternatives definition

Step 2.1: Superstructure generation

Action 2.1.1: A database search in the Biorefinery database reveals microalgae as available raw material and alternatives data from the previous work by Rizwan et al. [91] is stored in the database. The desired product name, biodiesel, has been used as argument for the search. Note that even though only one product is specified, the search returns the complete superstructure including possible co-products.

Action 2.1.2: The superstructure is now expanded with two additional raw materials: palm oil and waste cooking palm oil and alternative routes derived from previous work by Mansouri et al. [92] and Simasatitkul et al. [93]. The generation of a superstructure of alternatives for the production of biodiesel from palm oil and waste cooking palm oil is discussed below.

The PSIN representation has the ability to represent synthesis problems at different levels. That is, different types of superstructure can be posed with varying levels of aggregation and organizing the alternatives differently. Consider the example in Figure 9.2, which shows two alternative separation sequences for the downstream processing of biodiesel [89].

The representation of the two alternative sequences in Figure 9.2 in a superstructure is not trivial. The level of detail and organization in the superstructure depends on the problem that is being solved. First of all, the processing steps need to be determined and placed in order. A first approach to determining the sequence of processing steps could be based on the type of unit operation, hence obtaining four processing steps,

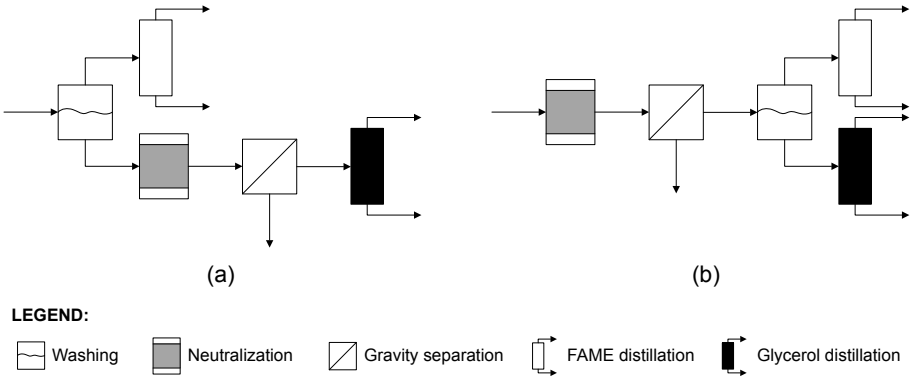


Figure 9.2. Two alternative downstream processing sequences for the biodiesel production process via transesterification of triglycerides: in (a) water washing is performed first, and neutralization of the catalyst in the aqueous phase is done afterwards, whereas in (b) neutralization is performed first and water washing is done afterwards.

namely washing, neutralization, gravity separation and distillation. The disadvantage of this approach is the difficulty in finding an order of the processing steps that fits all the alternative sequences. As illustrated in Figure 9.3, in the case considered here, the two alternative sequences differ in the order of the units, hence if the type of unit is taken as processing step, the superstructure representation will either have one or more repetitions of processing steps (Figure 9.3a) or a connection going from left to right without being a recycle (Figure 9.3b), which is not desirable, and even if some units are condensed in the same interval (Figure 9.3c) the problem might still appear.

Another possible criterion to organize the superstructure is using the order of the operations itself, hence the processing steps would be: first separation step, second separation step, third separation step, and so on (see Figure 9.4).

In this case, the first processing step of the sequence would be *separation 1*, and the alternatives would be performing separation between oil-phase and glycerol-phase first, or separating the catalyst first (via neutralization and gravity separation). Using this approach, various representations can be obtained: Figure 9.4a shows a superstructure based on separation steps where the final separation of the primary and secondary product streams are considered in the same processing step, whereas in Figure 9.4b they are placed in subsequent steps. In Figure 9.4c the two units for catalyst removal (neutralization and gravity separation) are considered together in the same processing interval. Finally Figure 9.4d is the most condensed version of the superstructure with only three steps since it the catalyst removal units are merged in one interval and the final separation of primary and secondary products are placed in one step. The choice of superstructure representation should be made based on the problem that is being

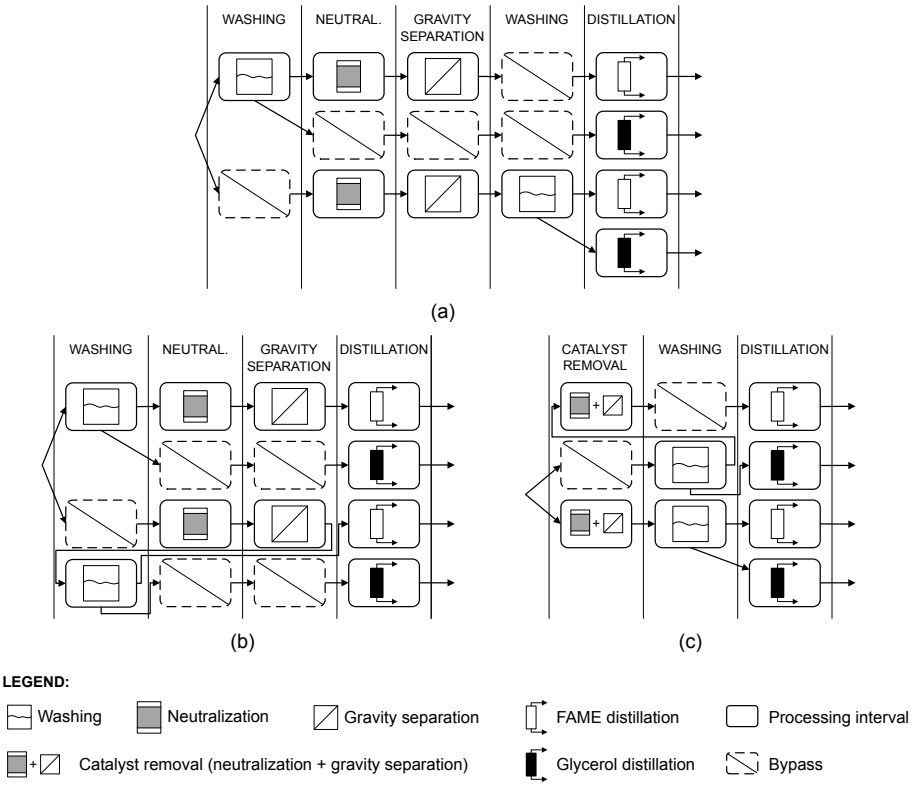


Figure 9.3. Alternative superstructure representations of three downstream processing sequences with processing steps based on type of equipment: in (a) washing appears twice as step to avoid connections from right to left, in (b) only one washing interval is used, but a connection from right to left appears, and in (c) two steps are merged, yet still obtaining a connection from right to left.

solved, for example for a problem where a unique neutralization reactor with multiple choice of gravity separation is considered, the versions in Figure 9.4c and Figure 9.4d are not convenient since by merging the two units in one interval, separate alternatives are not allowed.

Based on the previously mentioned considerations, the complete superstructure of alternatives for biodiesel production is built, which is shown in Figure 9.5. Processing steps and intervals are listed in Table 9.1 in Table 9.2, respectively.

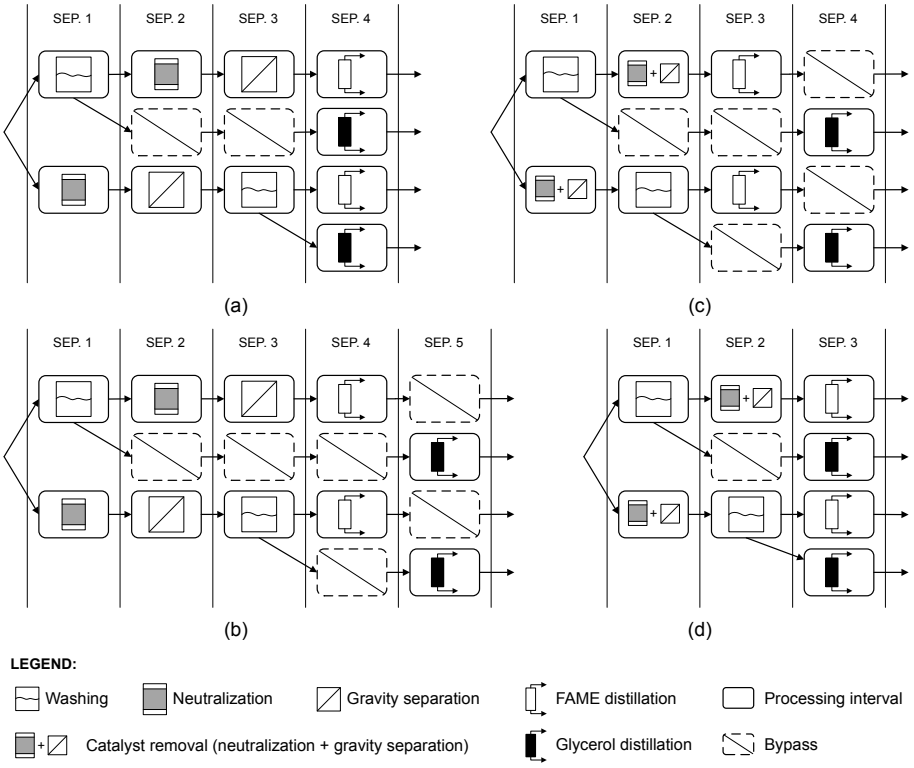


Figure 9.4. Alternative superstructure representations of two downstream processing sequences with processing steps based on sequence of separation: in (a) four processing steps are considered, whereas in (b) five are needed since the primary and secondary product purification are separated, in (c) this is maintained but two units are merged in represented together in one processing interval, neutralization and gravity separation, and in (d) the same is used as well as representing the final separation for biodiesel and glycerol in the same step. Note that with any of the variations of this approach all streams go from left to right.

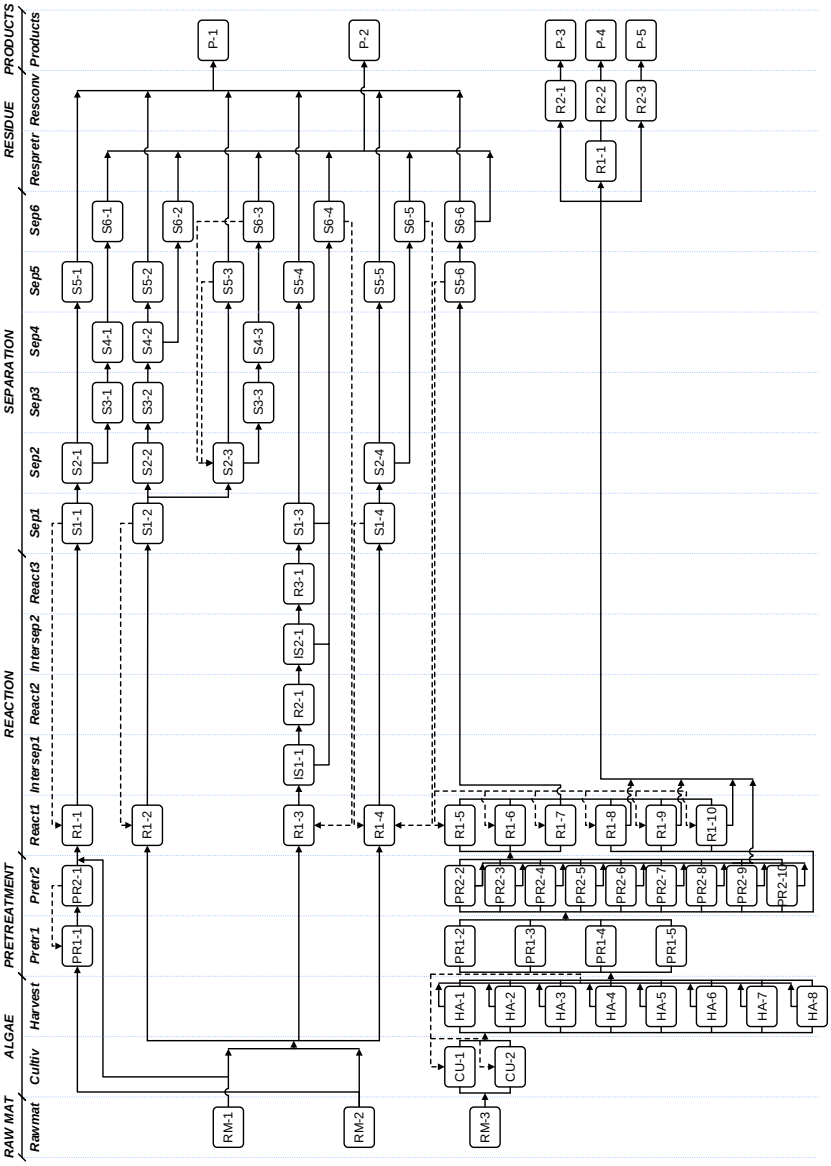


Figure 9.5. Biodiesel production superstructure (PSIN) of alternatives.

Table 9.1. List of processing steps in the superstructure for biodiesel production from various feedstocks.

Step	Description
RM	Raw materials
CU	Cultivation
HA	Harvesting
PR1	Pretreatment step 1
PR2	Pretreatment step 2
R1	Reaction step 1
IS1	Intermediate separation 1
R2	Reaction step 2
IS2	Intermediate separation 2
R3	Reaction step 3
S1	Downstream separation step 1
S2	Downstream separation step 2
S3	Downstream separation step 3
S4	Downstream separation step 4
S5	Downstream separation step 5
S6	Downstream separation step 6
RP	Residue pretreatment
RC	Residue conversion
P	Products

Table 9.2. List of processing intervals in the biodiesel superstructure.

Interval	Description	Reference
RM-1	Virgin palm oil	[89]
RM-2	Waste cooking palm oil	[89]
RM-3	Microalgae	[40, 94, 95, 96]
CU-1	Open pond system	[40, 97]
CU-2	Photobioreactor	[40, 97]
HA-1	Flocculation with poly electrolyte	[40, 98]
HA-2	Flocculation with NaOH	[40, 99]
HA-3	Flocculation with PGA	[40, 100]
HA-4	Flocculation with chitosan acid solution	[40]
HA-5	Bioflocculation + centrifugation	[40]
HA-6	Centrifugation	[40]

continued on next page

Table 9.2 – continued from previous page

Interval	Description	Reference
HA-7	Auto flocculation (induced by high pH)	[40]
HA-8	Microfiltration + centrifugation	[40]
PR1-1	Acid-catalyzed esterification	[89]
PR1-2	Grinding in liquid nitrogen	[40]
PR1-3	Drying + ultrasound	[40]
PR1-4	Drying + grinding + microwave + ultrasound	[40]
PR1-5	Drying	[40]
PR2-1	Glycerine washing and methanol recovery	[89]
PR2-2	Grinding-assisted lipid extraction	[40]
PR2-3	Ultrasound-assisted extraction by [Bmim][$MeSO_4$]	[40]
PR2-4	Ultrasound- and microwave-assisted lipid extraction	[40]
PR2-5	Wet lipid extraction	[40]
PR2-6	Solvent extraction (Bligh and Dyer's method)	[40]
PR2-7	Solvent extraction (Modified Bligh and Dyer's Method)	[40]
PR2-8	Supercritical fluid extraction	[40]
PR2-9	Extraction by ionic liquids mixture	[40]
PR2-10	Extraction by [Bmim][$MeSO_5$]	[40]
R1-1	Alkali-catalyzed transesterification	[89]
R1-2	Acid-catalyzed transesterification	[89]
R1-3	Solvent-free enzyme-catalyzed transesterification (1/3)	[101]
R1-4	Co-solvent enzyme-catalyzed transesterification	[101]
R1-5	Base catalyzed transesterification	[40]
R1-6	Acid catalyzed transesterification	[40]
R1-7	Enzymatic transesterification	[40]
R1-8	Alkaline in-situ transesterification	[40]
R1-9	Acidic in-situ transesterification	[40]
R1-10	Enzymatic in-situ transesterification	[40]
IS1-1	Removal of aqueous phase	[101]
R2-1	Solvent-free enzyme-catalyzed transesterification (2/3)	[101]
IS2-1	Removal of aqueous phase	[101]
R3-1	Solvent-free enzyme-catalyzed transesterification (3/3)	[101]

continued on next page

Table 9.2 – continued from previous page

Interval	Description	Reference
S1-1	Methanol recovery via distillation + recycle	[89]
S1-2	Methanol recovery via distillation + recycle	[89]
S1-3	Removal of aqueous phase in decanter	[101]
S1-4	Methanol recovery via distillation + recycle	[101]
S2-1	Aqueous phase separation from oil-phase	[89]
S2-2	Neutralization	[89]
S2-3	Phase separation	[89]
S2-4	Phase separation	[101]
S3-1	Neutralization	[89]
S3-2	Solids removal	[89]
S3-3	Neutralization	[89]
S4-1	Solids removal	[89]
S4-2	Phase separation	[89]
S4-3	Solids removal	[89]
S5-1	FAME purification 1	[89]
S5-2	FAME purification 2	[89]
S5-3	FAME purification 3	[89]
S5-4	FAME purification 4	[101]
S5-5	FAME purification 5	[101]
S5-6	FAME purification 6	[40]
S6-1	Glycerol purification 1	[89]
S6-2	Glycerol purification 2	[89]
S6-3	Glycerol purification 3	[89]
S6-4	Glycerol purification 4	[101]
S6-5	Glycerol purification 5	[101]
S6-6	Glycerol purification 6	[40]
RP-1	Enzymatic hydrolysis	[40]
RC-1	Fast pyrolysis	[40]
RC-2	Fermentation	[40]
RC-3	Anaerobic digestion	[40]
P-1	Biodiesel	

continued on next page

Table 9.2 – continued from previous page

Interval	Description	Reference
P-2	Glycerol	
P-3	Bio-oil	
P-4	Bioethanol	
P-5	Biogas	

Step 2.2: Collection and storage of relevant data

Action 2.2.1: Data related to the microalgae alternatives, which have been retrieved from the database, are available in the database in a form that can directly be input in Super-O or an input file to solve the optimization problem along with the generic model file.

Action 2.2.2: For new alternatives used to expand the superstructure, data is collected from literature sources.

Action 2.2.4: All collected data is input into Super-O for the problem solution.

Action 2.2.5: Data related to new alternatives in the superstructure is added to the Biorefinery database, expanding the knowledge contained in it for future reuse. The following assumption is used throughout the case study: only triglycerides and free fatty acids are considered, hence no mono- and di-glycerides are taken into account, due to the lack of reaction data for the step-wise reaction.

9.2.3 Step 3: Modeling of single-location problem

Step 3.1: Single-location mathematical model setup

Action 3.1.1: Three different objective functions are required for this problem, one for each scenario. This can be achieved by either creating three model files with different objective functions, or by keeping the default model file and changing data parameter values in order to achieve the same effect. For example, for a scenario where GOI is the objective function, capital cost does not need to be considered. In this case, capital cost functions can be set to zero in Super-O, hence the capital cost term becomes to zero and EBIT reduces to GOI. Note that EBIT is the default objective function in Super-O.

Action 3.1.2: All model blocks corresponding to the basic generic model are used in this example. Location and transportation are not considered.

Step 3.2: Definition of scenario-based model inputs

Action 3.2.1: For all scenarios, an inlet flow rate of raw material of 10,000 kg/h (*i.e.*, 80 kt/y) is fixed. Moreover, a project lifetime of 80,000 h (*i.e.*, 10 y) is specified for the distribution of capital costs. As mentioned in Action 3.1.1, input data can be modified to change the objective function form, or this can be done from the model file.

Action 3.2.2: Input data (feedstock flow rate) is included in Super-O.

9.2.4 Step 4: Single-location solution

Step 4.1: Input file generation and problem solution

Action 4.1.1: The optimization problem is solved using CPLEX in GAMS, the generic model and the data from the database regarding the superstructure of alternatives.

Action 4.1.2: Problem statistics are given in Table 9.3. The optimal topology obtained for this problem is given in Table 9.4.

Table 9.3. Problem statistics.

Problem	NF	2
	NP	2
	NS	15
	NI	36
Model	NEQ	79119
	NV	76678
	NDV	78
	Problem type	MIP
Solution	Solver	CPLEX
	Execution time [s]	0.172

Table 9.4. Optimal topology for scenarios for biodiesel production.

Scenario	Topology
D-1	RM-2, PR1-1, PR2-1, R1-1, S1-1, S3-1, S4-1, S5-1, S6-1, P-1, P-2
D-2	RM-1, R1-1, S1-1, S3-1, S4-1, S5-1, S6-1, P-1, P-2
D-3	RM-1, R1-1, S1-1, S3-1, S4-1, S5-1, S6-1, P-1, P-2

Action 4.1.3 Continue to Step 8.

9.3 Results

The selected topologies for each scenario have been highlighted in Table 9.4 and depicted in Figures 9.6 and 9.7. Moreover, the objective function break down for each of them is displayed in Figure 9.8. It is observed that when GOI is used as objective function and utilities are not considered (scenario D-1), waste palm oil is selected as raw material to produce biodiesel and glycerol via acid-catalyzed esterification pretreatment followed by alkali-catalyzed transesterification. As utility costs are included (scenario D-2), the selected raw material changes to virgin palm oil, which is transesterified using the same method. In this second case, pretreatment is not necessary since waste oil is not used. The topology remains the same when the performance criterion is changed to EBIT (scenario D-3). In this case, accounting for equipment costs does have an effect on the objective function value, however it does not affect the selected topology. It should be mentioned that the capital costs are distributed over a 10 year period and only equipment purchase costs are included.

As highlighted in Chapter 2, the presented methodology applied in this example corresponds to the synthesis stage (the first stage in the three-stage approach). The outputs of this stage are not only the selected raw materials, technologies and products, but also the calculated values of variables related to the technologies which allow for

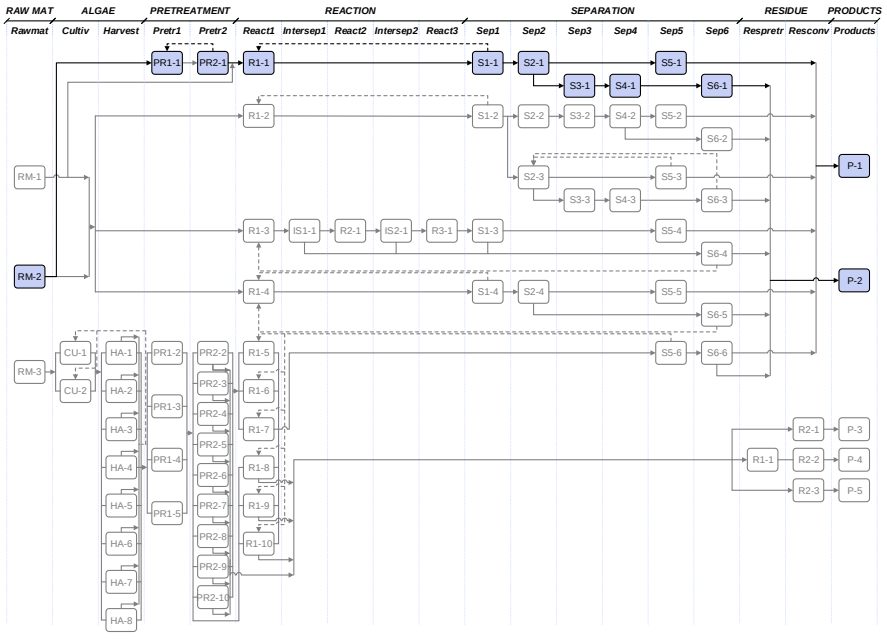


Figure 9.6. Biodiesel scenario D-1 selected topology.

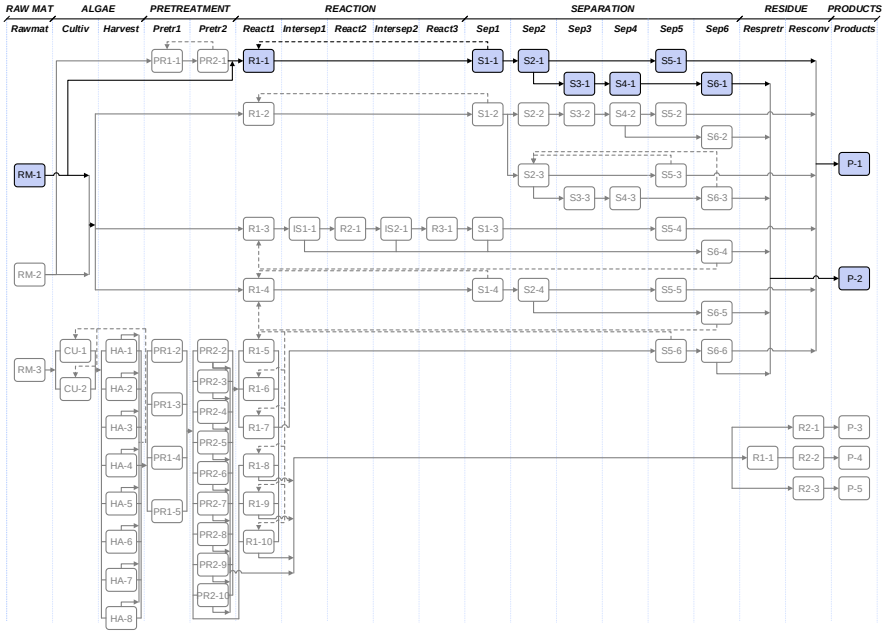


Figure 9.7. Biodiesel scenario D-2 and D-3 selected topology.

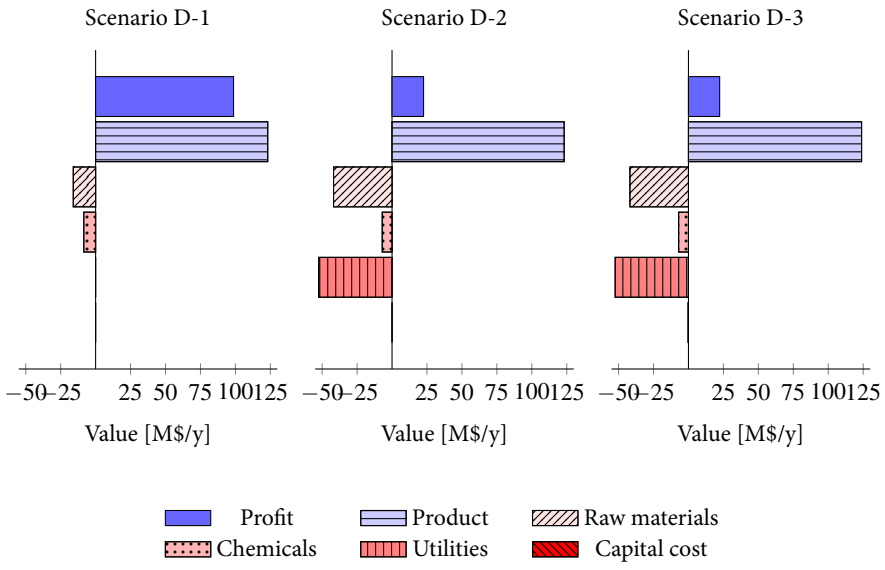


Figure 9.8. Objective function breakdown for scenarios D-1, D-2, and D-3.

further analysis even prior to further simulation or manipulation of the results. For example, the simple mass balance data is obtained, which is given in Tables 9.5 and 9.6 for scenarios D-1, and D-2 and D-3, respectively.

Table 9.5. Stream table scenario D-1.

	RM-2	PR1-1	PR2-1	R1-1	S1-1	S2-1
triolein	9400.00	9400.00	9306.00	465.30	465.30	465.30
methanol		107.64	101.16	3001.88	3001.88	180.11
moleate		629.80	629.80	9510.76	9510.76	9510.76
glycerol				919.50	919.50	919.50
water		38.27	4.59	4.59	4.59	4.59
NaOH		93.43	93.43	186.40	186.40	186.40
	S3-1	S4-1	S5-1	S6-1	P-1	P-2
methanol	99.01	99.01		16.18		16.18
moleate	11.30	11.30	9462.13	11.30	9462.13	11.30
glycerol	919.50	919.50		919.50		919.50
water	88.11	88.11	0.05	52.73	0.05	52.73

Table 9.6. Stream table scenarios D-2 and D-3.

	RM-1	R1-1	S1-1	S2-1	S3-1
triolein	10000.00	500.00	500.00	500.00	
methanol		1414.09	1414.09	84.85	46.64
moleate		9543.27	9543.27	9543.27	11.34
glycerol		988.08	988.08	988.08	988.08
water					45.00
NaOH		99.90	99.90	99.90	
H3PO4					0.68
Na3PO4					136.49
	S4-1	S5-1	S6-1	P-1	P-2
methanol	46.64		7.62		7.62
moleate	11.34	9494.47	11.34	9494.47	11.34
glycerol	988.08		988.08		988.08
water	45.00		26.93		26.93

In this case, location has not been considered as a variable in the optimization, however, a certain location-dependency is intrinsic to the problem given that data for alternatives connected to microalgae correspond to South Korea, whereas data for palm oil alternatives correspond to Malaysia.

It is worth mentioning that none of the alternatives concerning the use of microalgae appear in the optimal solutions. This is expected given that microalgae biorefineries have been reported not to be profitable solutions and technologies involved in them should be further improved to turn profitability around [40]. The fact that location constraints are not considered allows the optimizer to select the most profitable solution regardless of availability/demand constraints. However, in case that the problem was location-constrained, the lack of availability of certain feedstocks may favor less profitable solutions. It is in this context that the development of microalgal biorefineries has interest. Geographical locations with low biomass availability can profit from such alternative solutions.

Regarding biodiesel production from microalgae in South Korea, as pointed out by Rizwan et al. [91], all alternative routes lead to a negative objective function value, which indicates that the production of biodiesel from microalgae in South Korea, according to the current data and available technologies, is not economically feasible. The aforementioned conclusion provides little motivation towards moving on to the second stage (within the three-stage approach) and performing detailed design and analysis of the selected alternative(s). However, in order to direct further process developments, a systematic analysis of the results obtained from the optimization problem can help target improvements. This can be done with the model setup and results.

9.4 Conclusion

This example illustrates the application of the framework to a multi-product multi-feedstock case for the production of biodiesel. The developed framework is able to cope with the problem complexity in terms of number of alternatives and recycles. Moreover, even though location is not directly selected in the optimization, location issues are embedded in the problem, hence giving rise to the need of including them in the model and solution approach.

In this example, data retrieval from the Biorefinery Synthesis Database is perform, which contains data from previously solved synthesis problems and known biorefinery processing routes. The search and retrieval of data for this problem led to the automatic specification of 39 processing intervals and interconnections. It should be noted that once data is searched and retrieved from the database for a given problem, it can still be edited and/or modified through Super-O, in order to fit the specific problem needs or to account for changes that have taken place from the time when it was stored in the database (*e.g.*, pricing and market changes).

CHAPTER 10

Location-dependent synthesis of feedstock and process networks

– Ethanol production

The type, characteristics and availability of biomass-based feedstocks in different geographic locations are not homogeneous, which makes the problem of designing biofuel product processes a location-dependent problem.

This example represents a location-dependent synthesis of feedstock and process networks, where multiple feedstocks are considered for the production of a desired product, and location is included as a model output. Therefore, decisions on product, process and facility location selection are performed simultaneously. This allows to account for interactions between these decision layers. Moreover, the centralized vs distributed configuration of the processing network does not need to be pre-defined, as the configuration is selected as part of the solution. Transportation costs need therefore to be accounted for when considering distributed configurations.

10.1 Motivation for the production of ethanol

Ethanol is an attractive biofuel that can be used in blends with gasoline, thus requiring little or no modification to existing internal combustion engines. It is therefore a potential alternative for reducing emissions and fossil fuel dependency and various countries have implemented policies where it is required to use a minimum percentage of ethanol from renewable sources in blends with gasoline [102].

10.2 Formulation and solution

In the synthesis stage, different scenarios are considered with the objective of obtaining the optimal process topology for the production of ethanol from biomass by considering an array of renewable feedstocks and a list of geographical locations.

10.2.1 Step 1: Problem definition

The desired product is specified as fuel grade ethanol. Various alternatives in terms of biomass-based feedstocks and locations should be considered. The objective is to get an overview in terms of the most favorable process topology and feedstock in each of the considered locations as well as to determine the most suitable location-feedstock-process combination.

The following questions are relevant for this example:

Q1. What is the optimal feedstock and processing route for a given product?
Q2. Is the solution location dependent?
Q3. What is the best location-dependent solution?
Q4. How do transportation costs influence the selection of centralized vs distributed configuration?
Q5. What is the optimal feedstock-process-location combination considering transportation costs within the supply chain?

Step 1.1: Objective and scope definition

Action 1.1.1: Both synthesis and location-dependent synthesis problems are of interest. In this example, the problem objective is the maximization of profit (EBIT).

Action 1.1.2: Lignocellulosic raw materials are considered in this example, moreover, transformations are limited to biochemical (fermentation) conversion. Locations are considered at a global scale.

Step 1.2: Solution scenarios definition

Action 1.2.1: Location-dependent solutions need to be accounted for in location-dependent cases. The following will be considered: (i) synthesis problem for a series of locations to test the location dependency of the solution; (ii) location-dependent synthesis without transportation; (iii) location-dependent synthesis including transportation.

Action 1.2.2: An overview of the scenarios is shown in Figure 10.1.

The following scenarios are defined:

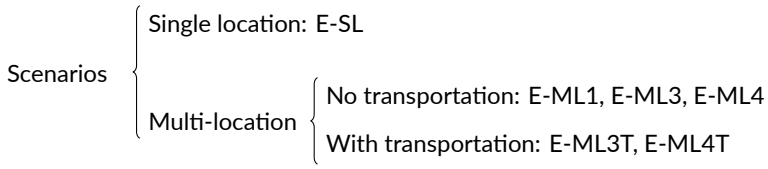


Figure 10.1. Overview of scenarios for the ethanol production example.

E-SL. synthesis problem for each single location to test the location dependency of the solution

E-ML1. location-dependent synthesis with single location constraint (raw materials, processing and product sale in a single location)

E-ML3. location-dependent synthesis with three location sections (sourcing, processing, sale), no transportation

E-ML3T. location-dependent synthesis with three location sections (sourcing, processing, sale), including transportation

E-ML4. location-dependent synthesis with four location sections (sourcing, pretreatment, processing, sale), including transportation

E-ML4T. location-dependent synthesis with four location sections (sourcing, pretreatment, processing, sale), including transportation, fixed processing location US

The above-mentioned scenarios may represent more that one solution since, for the sake of simplicity, small variations of parameters have been considered as being part of a single scenario. For example, if a problem is solved for two values of product demand or transportation price, the results are given within the same scenario. This way the number of scenarios is manageable. These variations are indicated within each scenario and highlighted along with the results in Section 10.3.

10.2.2 Step 2: Alternatives definition

A superstructure of raw material and technology alternatives for the production of ethanol is generated.

Step 2.1: Superstructure generation

Action 2.1.1: According to the performed literature review, the following main processing steps are considered:

1. Handling of raw material 2. Pre-treatment 3. Hydrolysis 4. Fermentation 5. Recovery 6. Purification 7. Product

Overall, six biomass-based feedstocks are considered: wheat straw (WS), corn stover (CS), sugarcane bagasse (SB), switch grass (SG), hardwood chips (HWC), and cassava rhizome (CR). These have been found to be available in seven geographic locations: Brazil (BR), Canada (CA), China (CN), India (IN), Mexico (MX), Thailand (TH), and United States (US). The complete list of the processing steps involved in this example is given in Table 10.1.

Table 10.1. List of processing steps in the superstructure for ethanol production.

Step	Description
RM	Raw materials
HAND	Feedstock handling
PRE	Pretreatment
HYD	Hydrolysis
FERM	Fermentation
BIOR	Biomass removal
SEP1	Recovery step
SEP2	Purification step
PROD	Products

The raw material step (RM) contains available lignocellulosic raw materials, the handling step (HAND) refers to unloading and washing of biomass, which is followed by pretreatment (PRE). Next, hydrolysis (HYD) is performed (unless it is done in the same step as fermentation), followed by fermentation (FERM). The first step after fermentation is the removal of biomass (BIOR), which is a solid-liquid separation operation. Separation steps include the first recovery step and a final purification step, which may contain more than one operation in order to reach the desired purity, which is past the ethanol-water azeotropic point. Finally, the product (PROD) ethanol is obtained.

Technological alternatives for each of the steps are gathered to generate a superstructure of alternatives. Both separate hydrolysis-fermentation and simultaneous saccharification-fermentation are considered. Moreover, alternatives for purification to fuel grade ethanol are considered. It should be noted that these are represented by a single interval each in the last separation step, regardless of the real number of units involved (for example, distillation followed by membrane separation is represented as one interval). The generated PSIN is depicted in Figure 10.2, the intervals in the superstructure are listed in Table 10.2.

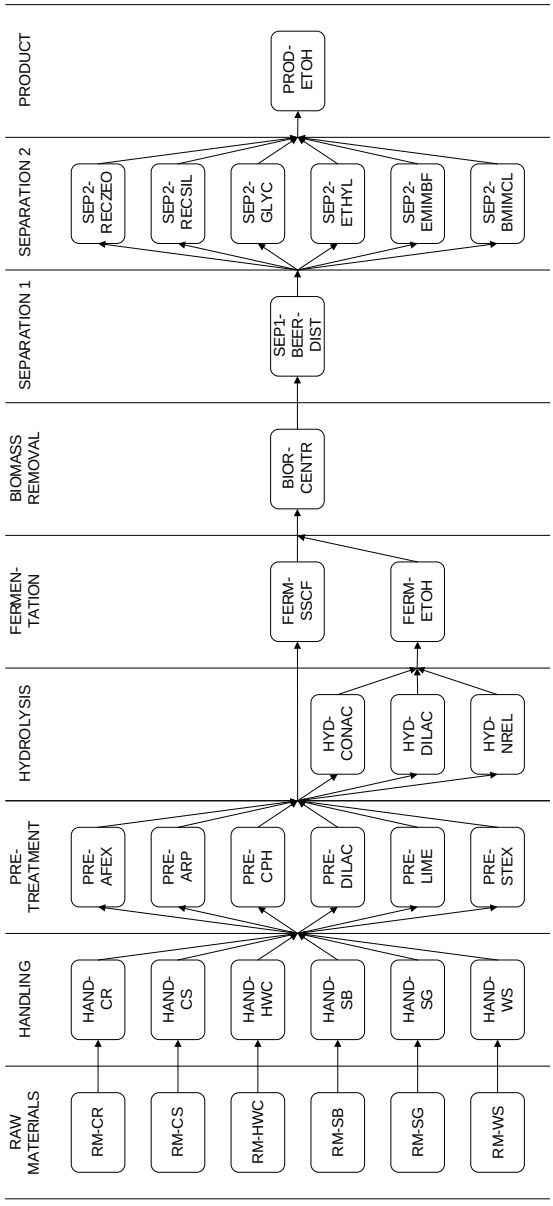


Figure 10.2. PSIN for the production of fuel-grade ethanol from various lignocellulosic feedstocks.

Table 10.2. Processing intervals for the production of ethanol from lignocellulosic biomass-based feedstocks.

Interval	Description
RM-CR	Cassava rhizome
RM-CS	Corn stover
RM-HWC	Hardwood chips
RM-SB	Sugarcane bagasse
RM-SG	Switch grass
RM-WS	Wheat straw
HAND-CR	Handling of cassava rhizome
HAND-CS	Handling of corn stover
HAND-HWC	Handling of hardwood chips
HAND-SB	Handling of sugarcane bagasse
HAND-SG	Handling of switch grass
HAND-WS	Handling of wheat straw
PRE-AFEX	Ammonia fiber explosion pretreatment
PRE-ARP	Ammonia recycle percolation pretreatment
PRE-CPH	Controlled pH pretreatment
PRE-DILAC	Dilute acid pretreatment
PRE-LIME	Lime pretreatment
PRE-STEX	Steam explosion pretreatment
HYD-CA	Concentrated acid hydrolysis
HYD-DA	Dilute acid hydrolysis
HYD-NE	Enzymatic hydrolysis using NREL enzyme
FERM-ETOH	Ethanol fermentation
FERM-SSCF	Simultaneous hydrolysis and saccharification
BIOR-CENTR	Biomass removal via centrifugation
SEP1-BEERDIST	Beer distillation
SEP2-RECZEO	Rectification followed by zeolite membrane
SEP2-RECSIL	Rectification followed by silica membrane
SEP2-GLYC	Solvent-based extraction with glycerol
SEP2-ETHYL	Solvent-based extraction with ethylene glycol
SEP2-EMIMBF	Extraction with ionic liqud EMIMBF4
SEP2-BMIMCL	Extraction with ionic liquid BMIMCl
PROD-ETOH	Fuel-grade ethanol

Step 2.2: Collection and storage of relevant data

Action 2.2.2: Relevant data is collected in this step. Location-dependent data is collected so as to solve the single-location problem for each location. That is, feedstock availability and price, and product demand and price data is collected and listed in Tables 10.3 and 10.4.

10.2.3 Step 3: Modeling of single-location problem

The single-location synthesis problem formulation is used on a location-by-location basis to determine the location-dependency of the solution.

Step 3.1: Single-location mathematical model setup

Action 3.1.1: Profit (EBIT) is selected as objective function.

Action 3.1.2: The model blocks corresponding to the basic generic model are selected.

Step 3.2: Definition of scenario-based model inputs

Action 3.2.1: A feedstock flow rate of 700,000 t/y or a product flow rate of 140,000 t/y, depending on the scenario, are selected based on a previous study [103].

10.2.4 Step 4: Single-location solution

An optimization problem is solved for each scenario using the basic generic model defined in Chapter 5.

Step 4.1: Input file generation and problem solution

Action 4.1.1: The linear (MILP) formulation is used and therefore CPLEX is used as solver. An input file is generated for each geographical location, so as to solve each single-location problem.

10.2.5 Step 5: Location alternatives definition

From this step onward, location is included as a decision in the model, therefore, location alternatives are defined.

Step 5.1: Sections and location alternatives generation

Action 5.1.1: Based on feedstock availability and ethanol demand, the seven locations previously outlined are selected as location alternatives. The set of locations consists

Table 10.3. Location-dependent availability and price of raw materials in various locations.

Feedstock	Interval	Location[a]	Availability	Unit	Price	Unit
Cassava rhizome	RM-CR	BR	23250	kt/y	89.10	$/t
		CN	4660	kt/y	100.10	$/t
		IN	8140	kt/y	161.68	$/t
		MX	18	kt/y	226.50	$/t
		TH	30020	kt/y	64.80	$/t
Corn stover	RM-CS	BR	29300	kt/y	33.95	$/t
		CA	3000	kt/y	11.02	$/t
		CN	220000	kt/y	30.97	$/t
		IN	1100	kt/y	8.05	$/t
		MX	18400	kt/y	52.03	$/t
		US	200000	kt/y	38.58	$/t
Hardwood chips	RM-HWC	BR	2610	kt/y	52.03	$/t
		CA	10500	kt/y	35.05	$/t
		CN	16290	kt/y	52.03	$/t
		MX	97	kt/y	52.03	$/t
		TH	790	kt/y	52.03	$/t
		US	19100	kt/y	35.27	$/t
Sugarcane bagasse	RM-SB	BR	148000	kt/y	7.60	$/t
		CN	4750	kt/y	9.10	$/t
		IN	6400	kt/y	7.30	$/t
		MX	12600	kt/y	15.40	$/t
		TH	12000	kt/y	2.05	$/t
		US	4620	kt/y	35.00	$/t
Switch grass	RM-SG	CA	80200	kt/y	125.00	$/t
		US	560	kt/y	50.00	$/t
Wheat straw	RM-WS	CA	12300	kt/y	30.80	$/t
		CN	188000	kt/y	10.00	$/t
		IN	185400	kt/y	28.10	$/t
		MX	4610	kt/y	7.30	$/t
		TH	1	kt/y	47.20	$/t
		US	82400	kt/y	35.00	$/t

[a] BR: Brazil, CA: Canada, CN: China, IN: India, MX: Mexico, TH: Thailand, US: United States.

Table 10.4. Location-dependent values of demand and price of product in various locations.

Product	Interval	Location[a]	Demand	Unit	Price	Unit
Fuel grade ethanol	PROD-ETOH	BR	21185	kt/y	766.87	$/t
		CA	1302	kt/y	746.00	$/t
		CN	2428	kt/y	657.00	$/t
		IN	630	kt/y	769.00	$/t
		MX	5	kt/y	849.18	$/t
		TH	998	kt/y	769.00	$/t
		US	44221	kt/y	798.54	$/t

[a] BR: Brazil, CA: Canada, CN: China, IN: India, MX: Mexico, TH: Thailand, US: United States.

of seven geographical areas (countries): Brazil (BR), Canada (CA), China (CN), India (IN), Mexico (MX), Thailand (TH), and United States (US). Note that all seven locations are considered for feedstock sourcing, plant location selection and product market. This is shown in Figure 10.5.

Action 5.1.1: Various problem setups in terms of processing sections are considered in this example depending on the scenario. First, a single (optimal) location is investigated. However, a minimum of three location sections is defined since the first and last processing steps automatically become the first and last sections. Therefore, the single-location scenario is created through the addition of a constraint that ensures that the selected location for the three sections is the same. This corresponds to scenario E-ML1 (Ethanol-Multi Location constrained to 1 location), as shown in Figure 10.3.

Next, various processing sections are considered. The legend for symbols used in location-based synthesis scenarios is shown in Figure 10.4. The next scenario consists of single section for processing, additionally to the raw material and product sections. This corresponds to scenario E-ML3 (Ethanol-Multi Location with maximum 3 locations allowed). This can yield to different locations from section to section, hence making the incorporation of transportation costs relevant, which is considered in scenario E-ML3T. The ePSIN for scenarios E-ML3 and E-ML3T is depicted in Figure 10.5.

The next scenario is formulated so that the processing chain is divided into two sections (namely pretreatment and processing). This scenario is named E-ML4 (Ethanol-Multi Location with maximum 4 locations allowed). The variation of this scenario to include transportation costs is E-ML4T. The ePSIN representation for scenarios E-ML4 and E-ML4T is shown in Figure 10.6.

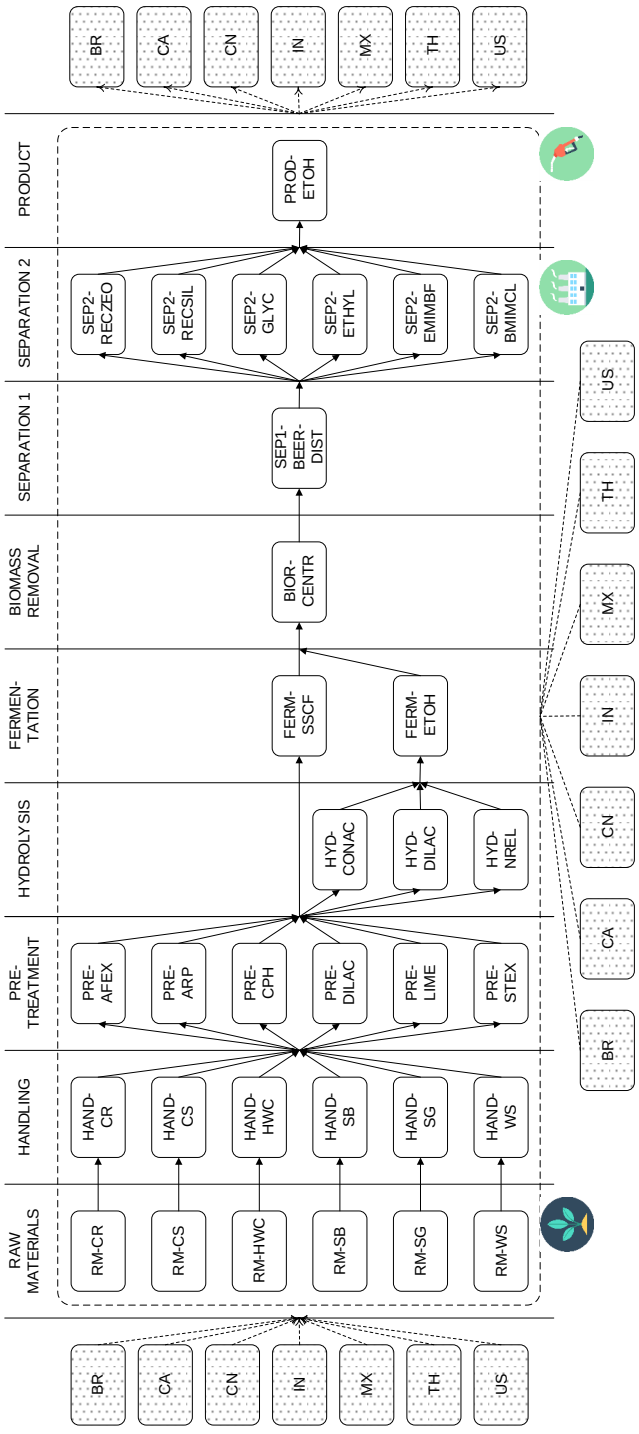


Figure 10.3. PSIN representation of scenario ML1.

Step 5.2: Collection and storage of location-based data

Action 5.2.2: Additional location-based data is collected and stored. Since the single-location problem was solved for various locations, some data is already available. Data obtained in this step includes supply (or availability) and demand data, and transportation prices. Note that a simple transportation model is used in this case study, with a single transportation mode and a single transportation price independent of location. Additionally, distances between geographical locations are determined. Assumptions are made with regard to distances between geographical locations: distances are calculated from and to a single point of each area, the shortest possible route is considered only, distances are not broken down into different transportation modes.

10.2.6 Step 6: Modeling of multi-location problem

Action 6.6.1: EBIT is selected as objective function, where transportation cost is included in operating costs for scenarios where data for it is provided (E-ML3T and ML4T).

Step 6.1: Multi-location mathematical model setup

The linear model blocks corresponding to the extended generic model are selected.

Action 6.1.1: EBIT is used as objective function for all location-based scenarios.

Action 6.1.2: The linear version of the extended generic model, as defined in Table 5.1 on page 61, is used in this example for location-dependent synthesis.

Action 6.1.3: Additional logic constraints are added to the model for some of the defined scenarios.

In scenario E-ML1, the location selection problem is solved, yet a single location should be selected for the entire value chain (raw material, process, product). The definition of a single process section enforces a single location for all processing intervals, except raw materials and products, which not only can be placed in a different location, but

Figure 10.4. Legend ethanol location-based scenarios.

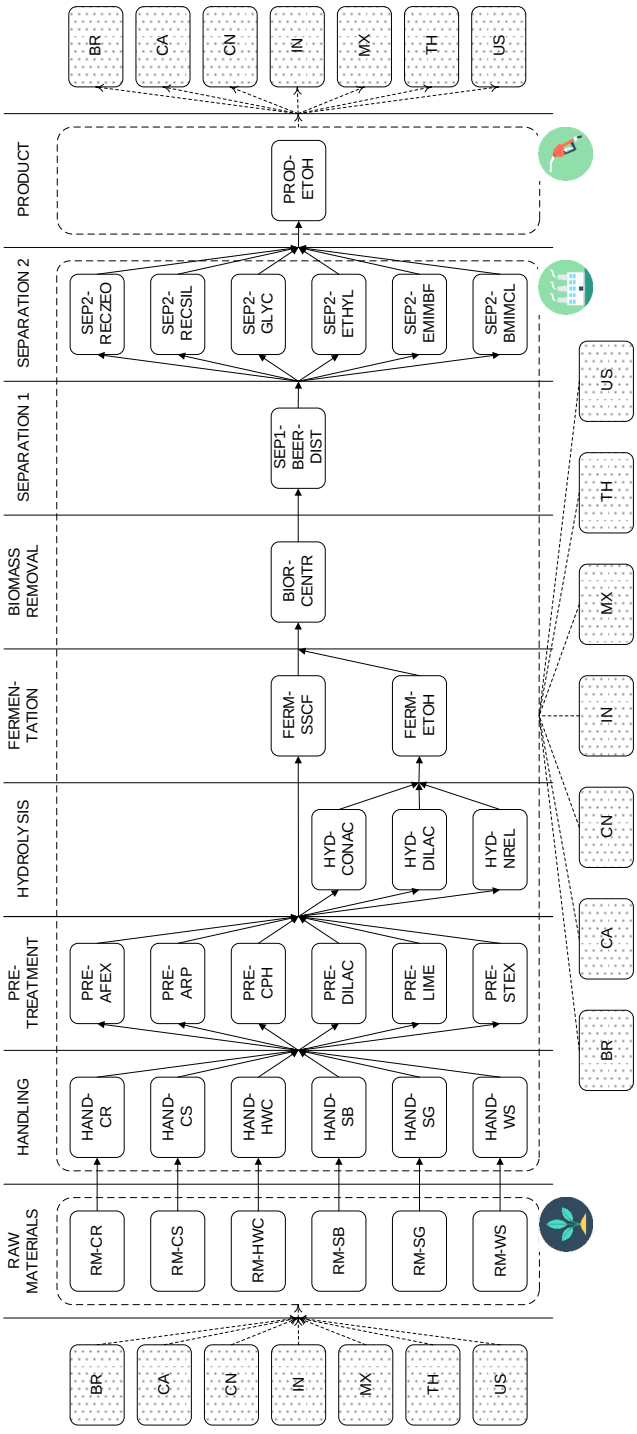


Figure 10.5. PSIN representation of scenarios ML3 and ML3T.

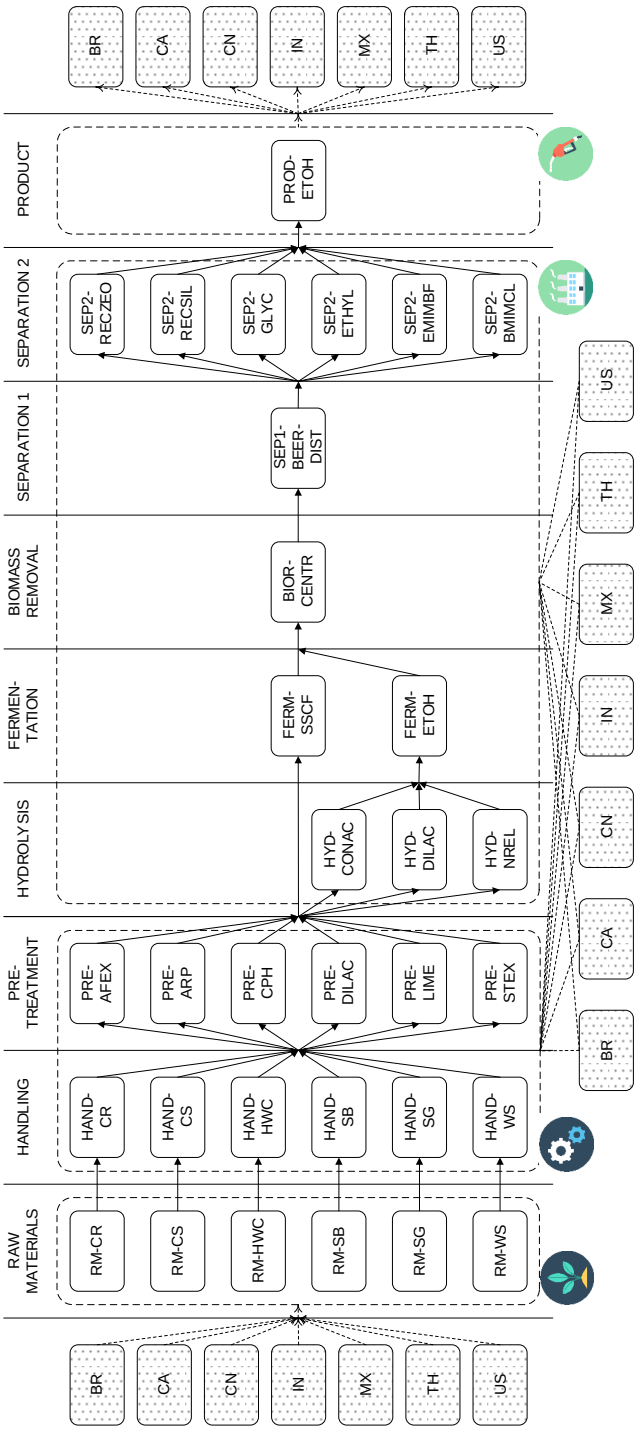


Figure 10.6. PSIN representation of scenarios ML4 and ML4T.

can be sourced from/distributed to multiple locations. First, for the single raw material case, the following constraint is added:

$$\sum_{k} \left(y^{k} \upsilon^{k,st} \right) \quad \forall st = 1 \tag{10.1}$$

A single location for each processing section is enforced with the following constraint:

$$\sum_{l} x^{t,l} = 1 \quad \forall l \tag{10.2}$$

A single location across processing sections is ensured as:

$$x^{t,l} - x^{tt,l} \leq 0 \quad \forall l, t, tt \tag{10.3}$$

Step 6.2: Definition of scenario-based model inputs

Action 6.2.1: **Scenario-based model inputs for this example are specified along with the results in Section 10.3.**

10.2.7 Step 7: Multi-location solution

Step 7.1: Input file generation and problem solution

Action 7.1.1: **CPLEX is selected to solve the MILP problems.**

10.3 Results

Results are summarized in this section in terms of the scenarios defined in Section 10.2.1. Variations of these scenarios are considered and defined as sub-scenarios, the details of which are outlined in the following sections. The legend used for graphical representation of the solutions is represented in Figure 10.7.

10.3.1 Single location solution

For each of the locations considered in this example, the single-location synthesis problem is solved to obtain the optimal process topology, selected feedstock and process flowrates.

Scenario E-BM

A benchmark solution run is performed to assess the accuracy of the model, with respect to detailed simulation results from a prior study for ethanol production from hardwood chips in United States via dilute acid pretreatment and simultaneous saccharification and fermentation (SSCF) [103]. Decision variables related to the process topology and feedstock selection are fixed to match the raw material and route used in the mentioned study. This way, simple mass and energy balance of the pre-defined process are performed, which can be used to compare to published results for the process.

Scenario E-SL

In the single-solution scenario (E-SL), the basic synthesis model is used to solve the ethanol synthesis problem for each geographical location. Therefore, seven sub-scenarios arise, one for each location. The objective of this part of the study is to understand the location dependency of biorefinery synthesis problems. It should therefore be noted that this was performed prior to some of the framework extensions presented in this thesis, leading to the motivation to perform said extensions.

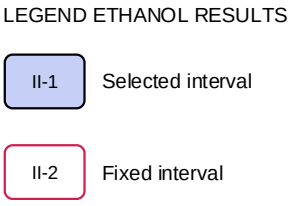


Figure 10.7. Legend ethanol location-based results.

The model formulation corresponds to the basic synthesis model where piecewise underestimators of the capital cost functions are used, hence leading to an MILP formulation, which is solved using CPLEX via GAMS.

Two variations of this scenario are considered. The first one (E-SLa) corresponds tot he maximization of the gross operating income (GOI), whereas the second one (E-SLb) includes capital costs, hence maximizing the earning before interests and tax (EBIT). Note that each sub-scenario is solved l times, where l is the number of locations, in this case seven. Location is therefore not an output of the model or an input variable per se, it determines the values of location-dependent data, such as prices. In this scenario, supply and demand constraints are not enforced directly, yet unavailable raw materials in a given location are not included in the set of alternatives for that location. It should be noted that the single-location cases of this example are constrained to a single raw material (i.e. feedstock mixtures are not allowed).

The results of maximizing GOI are given in Table 10.5 (top) and the maximization of EBIT yields the solutions listed in Table 10.5 (bottom). Problem statistics for the model and solution are listed in Table 10.6. Note that variations in the number of equations and problem size across locations responds to the difference in number of feedstocks available in each location.

The solutions outlined in Table 10.5 show, as expected, variations in the optimal solution across locations in terms of process structure, selected feedstock and profit value. This indicates synergies between the various decisions taken at the synthesis level. In terms of process structure, most differences are observed upstream, around the pretreatment step, which is expected given: (i) the larger number of alternatives upstream, and (ii) the higher non-homogeneity of material flows in the first steps of the superstructure. That is, as processing takes place, the material flows become more and more similar to each other, regardless of the initial raw material, hence decisions on their processing are less influenced by location. Nevertheless, some variations in the downstream structure are observed, which are due to geographical variations of energy and added chemicals prices (*e.g.*, solvents).

It is therefore clear that for this class of problems, simultaneous optimization of process topology, feedstock selection, and location is required. This is possible using the extended model for location-based synthesis, the results of which are shown and discussed in Section 10.3.2.

Table 10.5. Selected topology and feedstock for ethanol production scenarios: E-SLa, maximization of GOI (top), and E-SLb, maximization of EBIT (bottom).

Location	RM	HAND	PRE	HYD	FERM	BIOR	SEP1	SEP2	PROD	Profit
BR	HWC	HWC	ARP	–	SSCF	CENTR	BEER	RECTSIL	ETOH	9.70
CA	CS	CS	CPH	–	SSCF	CENTR	BEER	RECTSIL	ETOH	77.19
CN	HWC	HWC	CPH	–	SSCF	CENTR	BEER	RECTZEO	ETOH	11.86
IN	CS	CS	CPH	–	SSCF	CENTR	BEER	RECTSIL	ETOH	82.19
MX	SB	SB	CPH	DILAC	ETOH	CENTR	BEER	RECTSIL	ETOH	-1.80
TH	CR	CR	ARP	–	SSCF	CENTR	BEER	RECTSIL	ETOH	55.34
US	HWC	HWC	CPH	–	SSCF	CENTR	BEER	RECTZEO	ETOH	22.42

Location	RM	HAND	PRE	HYD	FERM	BIOR	SEP1	SEP2	PROD	Profit
BR	HWC	HWC	ARP	–	SSCF	CENTR	BEER	RECTSIL	ETOH	6.53
CA	CS	CS	CPH	–	SSCF	CENTR	BEER	RECTSIL	ETOH	74.41
CN	HWC	HWC	CPH	–	SSCF	CENTR	BEER	RECTZEO	ETOH	8.69
IN	CS	CS	CPH	–	SSCF	CENTR	BEER	RECTSIL	ETOH	76.41
MX	SB	SB	CPH	DILAC	ETOH	CENTR	BEER	RECTSIL	ETOH	-7.80
TH	CR	CR	ARP	–	SSCF	CENTR	BEER	RECTSIL	ETOH	46.96
US	HWC	HWC	CPH	–	SSCF	CENTR	BEER	RECTSIL	ETOH	19.25

Table 10.6. Problem, model and solution statistics for scenario E-SL of ethanol production.

		BR	CA	CN	IN
Problem	NL	1	1	1	1
	NF	3	4	4	4
	NP	1	1	1	1
	NS	9	9	9	9
	NI (excl. NF and NP)	29	29	29	29
Model and solver	NEQ	120,230	87,385	127,394	127,394
	NV	115,777	83,632	122,796	122,796
	NDV	64	54	66	66
	Problem type	MIP	MIP	MIP	MIP
	Solver	CPLEX	CPLEX	CPLEX	CPLEX
	Execution time [s]	0.609	0.453	0.64	0.639
		MX	TH	US	
Problem	NL	1	1	1	
	NF	4	5	5	
	NP	1	1	1	
	NS	9	9	9	
	NI (excl. NF and NP)	29	29	29	
Model and solver	NEQ	127,394	127,394	134,766	
	NV	122,796	122,796	130,023	
	NDV	66	66	68	
	Problem type	MIP	MIP	MIP	
	Solver	CPLEX	CPLEX	CPLEX	
	Execution time [s]	0.655	0.608	0.687	

10.3.2 Multi-location solution

Scenario E-ML1

In this scenario (ethanol multi-location constrained to 1 location, E-ML1a) the problem is constrained to a single location (that is, raw materials, process and product should be placed in a single location). Some variations of this scenario are considered:

ML1a. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material

ML1b. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to **100% of the real demand**, the solution is constrained to a single raw material

ML1c. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to **100% of the real availability**, product demand is set to 100% of the real demand, the solution is constrained to a single raw material

ML1d. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 100% of the real demand, **multiple raw materials** in each single location are allowed

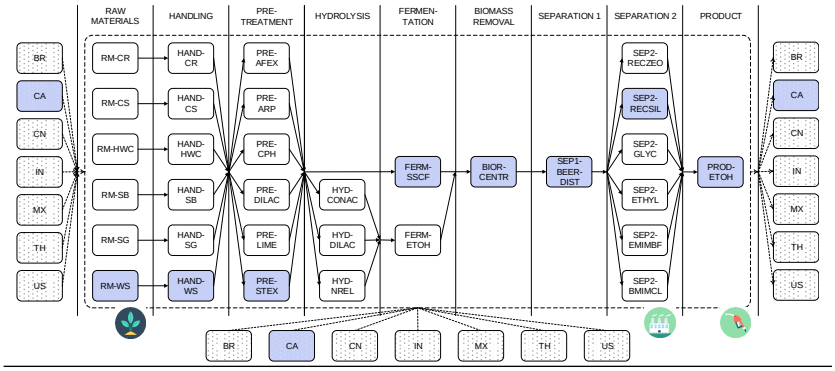
The objective of these variations is to ensure that the model accounts for supply/demand variations and multiple raw materials. The optimal location and route for scenario E-ML1a and the results associated with it are given in Table 10.7. It is observed that, in contrast with scenario E-SL, where India was the most profitable location, Canada is selected in this case. This is due to the supply and demand constraints, that allow a maximum of 10% of available raw material to be processed, and a maximum of 10% demand satisfied.

Additionally, integer cuts are implemented to explore the solution space. Near-optimal solutions for scenario E-ML1a are ranked in Table 10.8. These show the ability of the developed model of capturing synergies between decision levels (feedstock, processing route, location), hence confirming the importance of integrating these decisions.

Similarly, top-ranked solutions for scenario E-ML1b are given in Table 10.9, which having profit values lower than the previous scenario show that demand constraints that represent an upper bound on product flowrate into each given market have an effect on profitability. In a given market, these mathematical constraints represent the fact that only a smaller share of the market might be in need for additional supply, as the remaining might be covered by other firms or similar products. This upper bound

Table 10.7. Optimal solution for scenario E-ML1a.

Optimal solution scenario E-ML1



Optimal route

RM-WS, HAND-WS, PRET-STEX, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTZEO, PROD-ETOH

Feedstock			Product		
WS / CA	700.00	kt/y	ETOH / CA	140.00	kt/y

Term	Revenue		Cost		Objective function	
Product	90.98	M$/y				
Feedstock			6.10	M$/y		
Chemicals			25.50	M$/y		
Utilities			7.64	M$/y		
Transportation			–	–		
Capital			51.92	M$		
Profit (EBIT)					46.54	M$/y

might not be a known value and it is dynamic based on changes in market dynamics and competition. The solution show that accounting for this upper bound generates an alternative optimal solution not only in terms of plant location, but also for the selected feedstocks and processing route.

In scenario E-ML1c, an even less constrained case is considered, where all raw material in each region is considered available for processing and the entire demand can be covered in each single location. Results are listed in terms of top-ranked solutions in 10.10. This shows mainly effects on the selected raw material in some of the considered locations, which affects the processing route and profitability as well.

In scenarios E-ML1a to E-ML1c, the problem has been constrained to a single feed-

Table 10.8. Selected topology and location for top-ranked solutions for scenario E-ML1a. In this case, integer cuts on the x variables are used (x represents the location selection for each process section). Maximum availability is set to 10% of the total availability and maximum demand to 10% of the total demand in each location.

Solution rank	1	2	3	4	5
RM-HWC			1	1	
RM-SB		1			
RM-WS	1				1
HAND-HWC			1	1	
HAND-SB		1			
HAND-WS	1				1
PRET-AFEX					1
PRET-CPH		1	1	1	
PRET-STEX	1				
HYD-DILAC					1
FERM1-SSCF_1	1	1	1	1	
FERM2-SSCF_2	1	1	1	1	
FERM3-SSCF_3	1	1	1	1	
FERM1-ETOHFERM					1
BIOR-CENTR	1	1	1	1	1
SEP1-BEERDIST	1	1	1	1	1
SEP2-RECTSIL	1			1	
SEP2-RECTZEO		1	1		1
PROD-ETOH	1	1	1	1	1
Location RAW	CA	IN	TH	CN	US
Location PROC	CA	IN	TH	CN	US
Location PROD	CA	IN	TH	CN	US
$EBIT$ [M$/y]	46.54	45.36	19.25	8.69	-10.44
S_{PROD} [M$/y]	90.98	64.75	63.49	52.23	28.32
C_{RAW} [M$/y]	6.10	12.86	18.98	0.72	4.46
C_{CHEM} [M$/y]	25.50	23.59	19.84	8.65	18.58
C_{UTIL} [M$/y]	7.64	2.82	1.56	5.16	7.38
C_{CAP} [M$]	51.92	48.53	31.70	31.70	83.35

Table 10.9. Selected topology and location for top-ranked solutions for scenario E-ML1b. In this case, integer cuts on the $\boldsymbol{x}$ variables are used ($\boldsymbol{x}$ represents the location selection for each process section). Maximum availability is set to 10% of the total availability and maximum demand to 100% of the total demand in each location.

Solution rank	1	2	3	4	5
RM-CR			1		
RM-HWC				1	1
RM-WS	1	1			
HAND-CR			1		
HAND-HWC				1	1
HAND-WS	1	1			
PRET-ARP			1		
PRET-CPH	1	1		1	
PRET-STEX					1
FERM1-SSCF_1	1	1	1	1	1
FERM2-SSCF_2	1	1	1	1	1
FERM3-SSCF_3	1	1	1	1	1
BIOR-CENTR	1	1	1	1	1
SEP1-BEERDIST	1	1	1	1	1
SEP2-RECTSIL			1		1
SEP2-RECTZEO	1	1		1	
PROD-ETOH	1	1	1	1	1
Location RAW	IN	CA	TH	CN	US
Location PROC	IN	CA	TH	CN	US
Location PROD	IN	CA	TH	CN	US
$EBIT$ [M$/y]	74.45	69.63	49.96	8.69	4.20
S_{PROD} [M$/y]	105.22	102.07	117.86	63.90	52.23
C_{RAW} [M$/y]	4.46	6.10	41.59	18.98	12.86
C_{CHEM} [M$/y]	12.80	12.87	9.45	19.84	40.89
C_{UTIL} [M$/y]	6.91	6.87	11.48	1.56	3.47
C_{CAP} [M$]	66.02	66.02	83.79	31.70	24.89

Table 10.10. Selected topology and location for top-ranked solutions for scenario E-ML1c. In this case, integer cuts on the $\boldsymbol{x}$ variables are used ($\boldsymbol{x}$ represents the location selection for each process section). Maximum availability is set to 100% of the total availability and maximum demand to 100% of the total demand in each location.

Solution rank	1	2	3	4	5
RM-CR			1		
RM-CS	1	1			
RM-HWC				1	1
HAND-CR			1		
HAND-CS	1	1			
HAND-HWC				1	1
PRET-ARP			1		
PRET-CPH	1	1		1	1
FERM1-SSCF_1	1	1	1	1	1
FERM2-SSCF_2	1	1	1	1	1
FERM3-SSCF_3	1	1	1	1	1
BIOR-CENTR	1	1	1	1	1
SEP1-BEERDIST	1	1	1	1	1
SEP2-RECTSIL	1	1		1	
SEP2-RECTZEO			1		1
PROD-ETOH	1	1	1	1	1
Location RAW	IN	CA	TH	US	CN
Location PROC	IN	CA	TH	US	CN
Location PROD	IN	CA	TH	US	CN
$EBIT$ [M$/y]	76.41	71.41	46.96	19.25	8.69
S_{PROD} [M$/y]	106.68	100.49	117.86	63.49	52.23
C_{RAW} [M$/y]	4.79	6.56	41.59	12.86	18.96
C_{CHEM} [M$/y]	13.90	13.98	9.45	25.39	19.84
C_{UTIL} [M$/y]	5.80	5.76	11.48	2.82	1.56
C_{CAP} [M$]	57.87	57.87	83.79	31.70	31.70

stock per location. Scenario E-ML1d allows multiple feedstocks to be selected in each single location. The same supply and demand limits as in E-M1b are considered. The results in Table 10.11 show that multiple feedstocks are selected for one location, leading to more profitable solutions with respect to E-ML1b, which was constrained to single feedstocks. Moreover, the case that shows multiple feedstocks, namely India, selects a feedstock that has low availability, and in order to reach the desired flow rate, it selects an additional feedstock. The problem being linear, it is expected that single feedstocks are selected unless an upper bound is reached.

Table 10.11. Selected topology and location for top-ranked solutions for scenario E-ML1d. In this case, integer cuts on the $\boldsymbol{x}$ variables are used ($\boldsymbol{x}$ represents the location selection for each process section). Maximum availability is set to 10% of the total availability and maximum demand to 100% of the total demand in each location. Additionally, multiple feedstocks are allowed.

Solution rank	1	2	3	4	5
RM-CR			1		
RM-CS	1				
RM-HWC				1	1
RM-WS	1	1			
HAND-CR			1		
HAND-CS	1	1			
HAND-HWC				1	1
HAND-SB			1		
HAND-WS	1	1			
PRET-ARP			1		
PRET-CPH	1			1	1
PRET-STEX		1			
FERM1-SSCF_1	1	1	1	1	1
FERM2-SSCF_2	1	1	1	1	1
FERM3-SSCF_3	1	1	1	1	1
BIOR-CENTR	1	1	1	1	1
SEP1-BEERDIST	1	1	1	1	1
SEP2-RECTZEO	1	1	1	1	1
PROD-ETOH	1	1	1	1	1
Location RAW	IN	CA	TH	US	CN
Location PROC	IN	CA	TH	US	CN
Location PROD	IN	CA	TH	US	CN
$EBIT$ [M$/y]	74.76	69.63	49.96	19.25	8.69
S_{PROD} [M$/y]	105.45	102.07	117.86	63.49	52.23
C_{RAW} [M$/y]	4.51	6.10	41.59	12.86	18.98
C_{CHEM} [M$/y]	12.97	12.87	9.45	25.37	19.84
C_{UTIL} [M$/y]	6.74	6.87	11.48	2.82	1.56
C_{CAP} [M$]	64.78	66.02	83.79	31.70	31.70
RM-CR [kt/y]			700		
RM-CS [kt/y]	110	700			
RM-HWC [kt/y]				700	700
RM-WS [kt/y]	590				

Scenario E-ML3

In this scenario, the value chain is divided into three process sections, which can each be located in a different location. Sub-scenarios are defined as:

ML3a. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material

ML3b. The **total product flowrate is fixed to 140 kt/y**, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material

ML3c. The total product flowrate is fixed to 140 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, **multiple raw materials are allowed**, **more than one location per section is allowed** for feedstocks and product

As well as multiple locations, these explore the differences between specifying the feedstock flow rather versus the product flow rate. Since different raw material have different yields to ethanol, specifying the feedstock flow rate leads to varying amounts of ethanol being produce, depending on the selected feedstock. However, generally a desired amount of output is given. As pointed out in Chapter 4, the extension in the model allows the specification of either sources or sinks flow rate, hence this is tested.

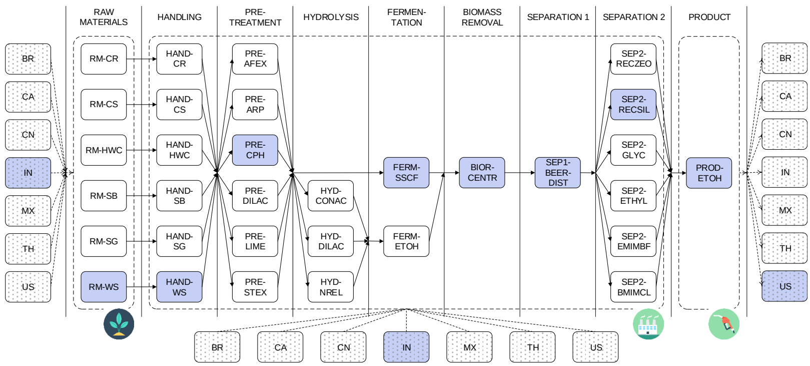
Scenario E-ML3a results are given in Table 10.12, which show that a geographically distributed network is favored. This is, however, not a realistic solution, due to the lack of transportation costs, which are included in scenario E-ML3T.

Scenario E-ML3b, the results of which are listed in Table 10.13, shows once more a very distributed processing network, given the lack of transportation costs in the objective function.

The last sub-scenario, E-ML3c, allows multiple feedstocks to be selected, instead of a single one, as well as raw materials and products to be sourced from or sold to multiple locations. Results shown in Table 10.14 show the selection of two raw materials from different locations and the sale of product in two markets as well, first a market is saturated and then the product is sold to another market.

Table 10.12. Results scenario E-ML3a.

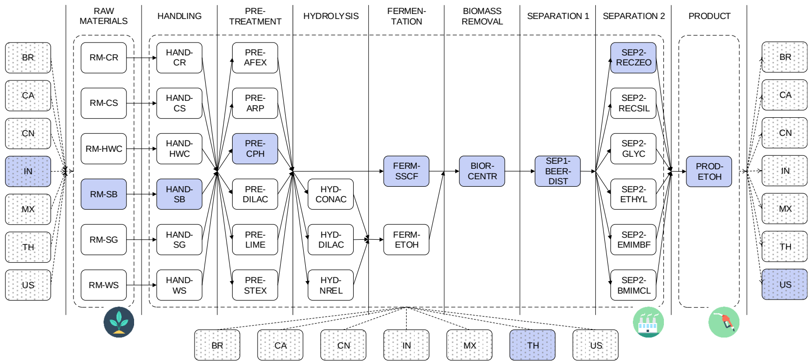
Optimal solution scenario E-ML3a					



Optimal route					
RM-WS, HAND-WS, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTSIL, PROD-ETOH					
Optimal locations					
Feedstock	Process			Product	
India	India			United States	
Feedstock			**Product**		
WS / IN	700.00	kt/y	ETOH / US	136.83	kt/y
Term	Revenue		Cost		Objective function
Product	109.29	M$/y			
Feedstock			4.46	M$/y	
Chemicals			12.80	M$/y	
Utilities			6.91	M$/y	
Transportation			–	–	
Capital			66.02	M$	
Profit (EBIT)					78.52 M$/y

Table 10.13. Results scenario E-ML3b.

Optimal solution scenario E-ML3b



Optimal route
RM-SB, HAND-SB, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTZEO, PROD-ETOH

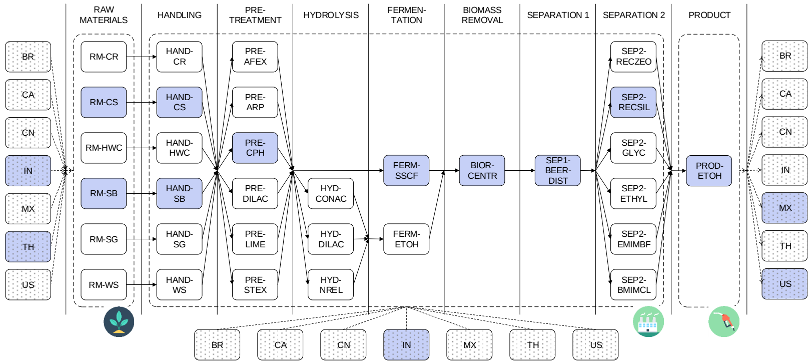
Optimal locations		
Feedstock	Process	Product
India	Thailand	United States

Feedstock			Product		
SB / IN	1,163.9	kt/y	ETOH / US	140.00	kt/y

Term	Revenue		Cost		Objective function	
Product	111.80	M$/y				
Feedstock			1.19	M$/y		
Chemicals			14.70	M$/y		
Utilities			7.82	M$/y		
Transportation			–	–		
Capital			71.35	M$		
Profit (EBIT)					80.95	M$/y

Table 10.14. Results scenario E-ML3c.

Optimal solution scenario E-ML3c							



Optimal route
RM-CS, RM-SB, HAND-CS, HAND-SB, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTSIL, PROD-ETOH

Optimal locations		
Feedstock	Process	Product
India	India	Mexico
Thailand		United States

Feedstock			Product		
CS / IN	1,100.0	kt/y	ETOH / MX	0.46	kt/y
SB / TH	982.68	kt/y	ETOH / US	139.54	kt/y

Term	Revenue		Cost		Objective function	
Product	111.82	M$/y				
Feedstock			1.76	M$/y		
Chemicals			14.60	M$/y		
Utilities			7.51	M$/y		
Transportation			–	–		
Capital			69.36	M$		
Profit (EBIT)					81.10	M$/y

Scenario E-ML3T

Given that geographically distributed solutions are explored in the previous scenarios, a more realistic approach is necessary, where transportation costs are accounted for. This is done in this scenario, ethanol production in a multi-location network with 3 processing sections including transport (E-ML3T). Three sub-scenarios considered are:

ML3Ta. The total feedstock flowrate is fixed to 700 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material, a transportation cost of 0.05 $/t/km is considered[104]

ML3Tb. The total **product flowrate is fixed to 140 kt/y**, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material, a transportation cost of 0.05 $/t/km is considered [104]

ML3Tc. The total product flowrate is fixed to 140 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, **multiple raw materials are allowed**, **multiple locations for each section are allowed**, a transportation cost of 0.05 $/t/km is considered [104]

The outcome of the first sub-scenario, E-ML3Ta, is summarized in Table 10.15, a distributed yet more centralized solution is observed. This solution is very close to that one of scenario E-ML3Tb (see Table 10.16), which only differs in that the product flow rate is specified, rather than the feedstock.

The multi-location case, scenario ML3Tc, in Table 10.17, presents a solution on the same lines as the previous sub-scenarios. However, in this case two raw materials are consumed, reaching the upper limit of one of them. This is a partly distributed solution in which the maximum available share of the local market is saturated first, and the reminder of product is transported to the nearest market. Even though the ethanol price is higher in United States than in Canada, sale in Canada is preferred in order to avoid transportation.

This scenario shows once more that synergies between decision levels should be accounted for from an early-stage of decision making.

Table 10.15. Results scenario E-ML3Ta.

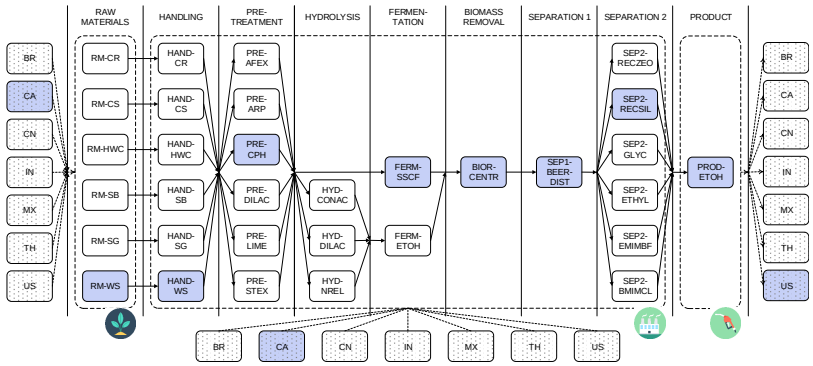
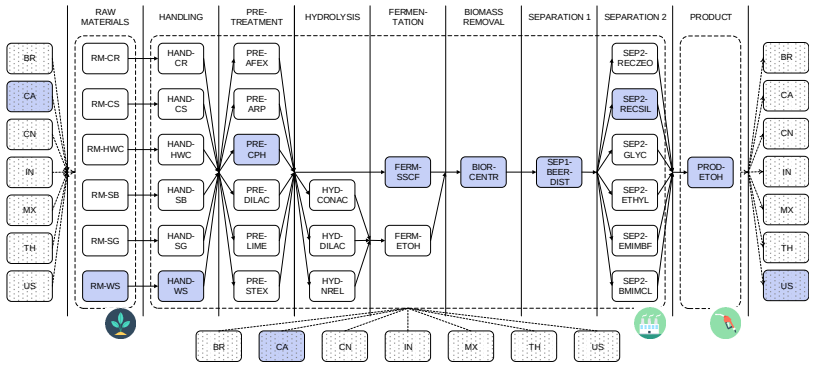
Optimal solution scenario E-ML3Ta						
						
Optimal route						
RM-WS, HAND-WS, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTSIL, PROD-ETOH						
Optimal locations						
Feedstock		Process			Product	
Canada		Canada			United States	
Feedstock			Product			
WS / CA	700.00	kt/y	ETOH / US		136.83	kt/y
Term	Revenue		Cost		Objective function	
Product	109.26	M$/y				
Feedstock			6.10	M$/y		
Chemicals			12.87	M$/y		
Utilities			6.87	M$/y		
Transportation			3.84	M$/y		
Capital			66.02	M$		
Profit (EBIT)					72.98	M$/y

Table 10.16. Results scenario E-ML3Tb.

Optimal solution scenario E-ML3Tb


Optimal route
RM-WS, HAND-WS, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTZEO, PROD-ETOH

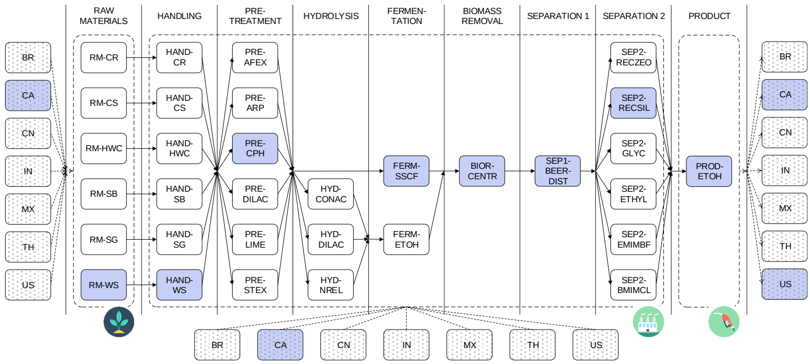
Optimal locations		
Feedstock	Process	Product
Canada	Canada	United States

Feedstock			Product		
WS / CA	716.22	kt/y	ETOH / US	140.00	kt/y

Term	Revenue		Cost		Objective function	
Product	111.80	M$/y				
Feedstock			6.24	M$/y		
Chemicals			13.17	M$/y		
Utilities			7.03	M$/y		
Transportation			3.93	M$/y		
Capital			67.21	M$		
Profit (EBIT)					74.70	M$/y

Table 10.17. Results scenario E-ML3Tc.

Optimal solution scenario E-ML3Tc



Optimal route

RM-CS, RM-WS, HAND-CS, HAND-WS, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTZEO, PROD-ETOH

Optimal locations

Feedstock	Process	Product
Canada	Canada	Canada
		United States

Feedstock			Product		
CS / CA	300.00	kt/y	ETOH / CA	130.00	kt/y
WS / CA	412.05	kt/y	ETOH / US	10.00	kt/y

Term	Revenue		Cost		Objective function	
Product	111.80	M$/y				
Feedstock			6.40	M$/y		
Chemicals			13.57	M$/y		
Utilities			6.51	M$/y		
Transportation			3.93	M$/y		
Capital			63.44	M$		
Profit (EBIT)					75.04	M$/y

Scenario E-ML4T

In this scenario, the question of centralized vs distributed network configurations is addressed. The definition of four process sections (raw material, pretreatment, process, and product) allows the individual selection of locations for each of them. The case including transportation is presented directly, since having a distributed system with no transportation does not make physical sense. The considered sub-scenarios are:

ML4Ta. The total product flowrate is fixed to 140 kt/y, feedstock availability is set to 10% of the real availability, product demand is set to 10% of the real demand, the solution is constrained to a single raw material, a transportation cost of 0.05 $/t/km is considered [104]

ML4Tb. A general overview of cases where the location of one of the processing sections is fixed and different values of transportation price are considered

Scenario E-ML4Ta yields the results in Table 10.18, which point towards the preference of a centralized network located in Canada with product distribution in United States. It it therefore observed that transportation of the product to a nearby location is preferred, given that the market price is higher. In this context, limitations of the transportation model implemented need to be highlighted. The transportation costs are calculated on a mass basis, instead of volume, and the same prices are used throughout the network for raw materials, intermediates, and products, which is a bold assumption. Therefore, it is expected that transportation, if present, is preferred towards intermediate or later sections of the process, given the lower mass flow rates.

For certain problems, it may be desired to explicitly define the location of one of the processing steps, and determine the optimal locations for the remaining steps. This type of problem could arise in the context of a processing location being available or desired, where the markets for feedstock and products need to be identified, and pretreatment can be performed on site or closer to the raw material collection location. These cases are explored in scenario E-ML4Tb for various locations (see Table 10.19). Moreover, the effect of transportation price changes is explored, in order to determine the influence of this parameter in the solution.

The results in Table 10.19 show that distributed networks tend to be preferred when the transporation price is not too high, that is, where pretreatment is close to the feedstock source, and processing takes place in another location. As transport price is increased, more centralized configurations are obviously preferred.

Table 10.18. Results scenario E-ML4Ta.

Optimal network scenario E-ML34Ta						
RM-WS, HAND-WS, PRET-CPH, FERM1-SSCF_1, FERM2-SSCF_2, FERM3-SSCF_3, BIOR-CENTR, SEP1-BEERDIST, SEP2-RECTSIL, PROD-ETOH						
Optimal locations						
Feedstock	Pretreatment		Process		Product	
Canada	Canada		Canada		United States	
Feedstock			Product			
WS / CA	716.22	kt/y	ETOH / US		140.00	kt/y
Term	Revenue		Cost		Objective function	
Product	111.80	M$/y				
Feedstock			6.25	M$/y		
Chemicals			13.17	M$/y		
Utilities			7.03	M$/y		
Transportation			3.93	M$/y		
Capital			67.21	M$		
Profit (EBIT)					74.70	M$/y

Table 10.19. Results of scenarios E-ML4Tb obtained by fixing the location of one section and varying transportation prices.

Scenario	Transport price		RM	Location				PROD	OF
	[$/t/km]	Type	Rate [kt/y]	RM	PRET	PROC	PROD	Rate [kt/y]	M$/y
b1	0.05	WS	700	CA	CA	US*	US	140	81.23
b2	0.5	WS	700	CA	CA	US *	US	140	75.38
b3	0.05	WS	700	IN	IN	TH *	US	140	85.53
b4	0.5	WS	700	IN	TH	TH *	TH / IN	40 / 100	57.83
b5	0.05	WS	700	CA	CA	BR *	US	140	70.58
b6	0.5	CR	700	BR	BR	BR *	US	140	-26.95
b7	0.05	WS	700	IN	IN	CN *	US	140	83.7
b8	0.5	WS	700	IN	IN	CN *	IN / TH	140	34.47
b9	5	WS	700	CN	CN	CN *	CN	140	13.91

* Fixed location

10.4 Conclusion

This example has been used to demonstrate the applicability and features of the framework in its full extent. First, single location solutions have been generated to motivate the need to include location selection as a model output. Then different variations of the location-dependent problem are explored, in which different features are tested: specifying feedstock flowrate vs product flowrate, defining multiple process sections, including transportation cost, and allowing multiple feedstocks, among others.

The results of this example show the capabilities of the framework in terms of assisting the synthesis of biorefineries providing a tool for systematic evaluation of alternatives including location selection and the decision of centralized vs distributed configuration.

The method and tools are therefore valid for a large range of problems containing the mentioned elements and concerning the highlighted decisions. Smaller geographical areas could be used for the solution of regional syntehsis of biorefineries. The results have also shown the need for further developing the transportation model to include multiple transportation modes, a more accurate representation of distances (for example, including a tortuosity parameter) and allowing the combination of various transportation modes to cover long distances. These enhancements would provide a more realistic numerical value of the transportation cost term, however, the model has been shown to be qualitatively correct.

PART IV

Conclusion

CHAPTER 11

Conclusion & future perspectives

11.1 Achievements

The design of biorefinery processing networks is a complex decision-making problem that can be formulated as a large MINLP. The complexity of this problem can be managed by dividing it into three sub-problems: synthesis, detailed design-analysis, and innovation. The solution approach for each of these sub-problems can be simultaneous or decomposition-based.

The synthesis of novel processing networks including biorefineries give rise to additional challenges in the early stages of design due to the location-dependency of their solution, the relatively low maturity level of these processes, and the fast changing political and social context around them. Research in this area is ongoing, however previous works generally consider a single biorefinery problem or a specific type of biorefineries, hence developing specific models for each given problem. Moreover, prior works normally focus on either process synthesis or supply chain aspects, missing the interconnection between decisions in these areas. Therefore, there is a clear need for the development a systematic framework for synthesis of biorefineries addressing these issues.

This has been addressed in this PhD project through the development of an integrated framework for the synthesis stage and its application to the synthesis of biorefineries. The framework consists of four key elements: a superstructure representation, a set of generic model blocks, solution strategies, and a data structure. The workflow consists of 8 steps, that cover the formulation and solution of synthesis and location-based synthesis of feedstock, process and product networks. The resulting framework is flexible, covering a wide range of problems, as has been discussed and shown through application examples. Moreover, it is able to cope with the necessary data for the solution of biorefinery synthesis problems, not only assisting in its systematization and storage, but also allowing its retrieval and reuse. The framework is generic, given that it can be applied not only to various biorefinery problems, but also to other application areas such as chemical process synthesis and synthesis of bio-processes. The software

implementation, Super-O, integrates the necessary methods and tools and allows an efficient use of the developed framework, hence reducing the time required to solve superstructure optimization problems. In addition, it provides access to this approach to less experienced users, with little prior knowledge of the modeling language and optimization procedures.

A framework for synthesis of processing networks was first proposed by Quaglia [71] alongside a software implementation named EOLO. This framework has been further developed and extended in this thesis, towards its application to a wider range of problems, including biorefineries. The main extensions are the location-dependency of the solution and a the Biorefinery Synthesis Database. The software tool Super-O is an updated version of EOLO. Issues related to data uncertainty were widely studied by Quaglia [71] and have therefore not been addressed in this project. The methods developed in this previous work can be directly applied in the context of biorefineries. Even though uncertainty studies are not described in this thesis, uncertainty information, when available, has been stored in the database.

As pointed out in the Introduction, relevant decisions that need to be made through the application of the framework include: the structure or topology of the process, selection of inputs (feedstocks) and outputs (products), the geographical configuration of the network, and aspects of the supply chain. All of these are addressed with the developed framework and shown through application examples. The conversion of sugarcane molasses is used to illustrate the features of the framework for the synthesis of process-product networks. The application to an example of biodiesel production shows the performance of the framework in a more complex and relevant example of feedstock-process-product synthesis. Location-dependency issues start to arise in this example and are addressed in the third example, concerning the production of ethanol. This example undertakes the production of ethanol from various lignocellulosic feedstocks in multiple locations, hence solving the location-based feedstock-process synthesis problem. First, single-location solutions are considered to test the location-dependency of the solution, which leads to including location selection and supply chain aspects in an extended model. The extended framework can provide location-dependent solutions for synthesis and address the issue of centralized vs distributed configurations.

11.2 Remaining challenges

Despite the value and scope of the developments achieved in this project, which have been highlighted in Section 11.1, some areas remain open for development.

Database

The developed data structure holds a value in itself and can be and has been implemented to other application areas, such as waste-water networks or carbon dioxide utilization. However, the Biorefinery Synthesis Database has been maintained and populated within this PhD project and needs to be further updated by including data for more technologies, feedstocks, products, and locations. Moreover, market data might become outdated and should be updated regularly.

Transportation

The need for a transportation model in the developed framework stems from the evaluation of distributed biorefinery networks. The model that is presented in this thesis can be further refined in order to obtain a more realistic representation of the transportation phenomenon. This can be done through allowing different transportation modes to be used to cover a distance (for example, railway followed by truck), accounting for tortuosity (to account for the fact that transportation generally does not follow straight lines), and adding the selection of transportation modes from different alternatives. Moreover, the definition of geographical locations and calculation of distances between them can be improved through the use of Geographic Information Systems (GIS).

Sensitivity

Integration of sensitivity analysis is a natural extension of the current framework, which would allow for identification of the main parameters affecting the solution. Moreover, including the range of validity of a certain solution for each parameter would be valuable as output of the framework.

Super-O

Super-O, the interface for formulating and solving synthesis problems, implements steps of the presented workflow and integrates the described tools. Additional efforts can be targeted towards developing this tool further, by creating a connection between the interface and database and expanding the library of problems that are used as examples to illustrate its application range.

11.3 Future perspectives

The following perspectives are suggestions and ideas for the development of biorefinery synthesis methods from a higher-level point of view.

Integration with other methods and tools

The value and power of the developed method would be further increased by establishing integration with available tools beyond the synthesis stage. For instance, a Computer-Aided Flowsheet Synthesis tool has been presented by Tula et al. [17], which automatically generates process superstructures based on thermodynamic insights given desired inputs and outputs and evaluate the performance alternatives using property data. This method has the capability of generating innovative alternatives and the means to evaluate them without prior knowledge of them, yet an enumeration approach is used for the solution, hence limiting problems to a manageable size. The integration of this tool and Super-O has potential since superstructures could be automatically generated, then data for known alternatives could be searched in the database and if not available, it could be estimated from property data. Then, the solution approach from Super-O could be used to obtain an optimal solution, being able to cope with larger problem sizes. Moreover, the method by Tula et al. [17] has connections with process simulators, that are used at the end of the synthesis stage, to automatically simulate the selected processing route.

Certificates of guarantee

Synthesis tools are powerful in that they provide means to evaluate many alternatives from an early stage when only limited data is available. As highlighted in this thesis, the outcomes of the synthesis stage (set of most promising alternatives) are not ready-to-build processes, but rather promising concepts that need to be further developed and analyzed using high fidelity models. The issues of operability, controlability, safety and dynamic performance of the developed process networks are generally not addressed in synthesis and need to be analyzed further on. However, they should possibly be addressed as early as possible in the design procedure. Parametric programming approaches can be used to obtain certificates of guarantee of the reliability of the obtained solutions [105].

Political and social context

The synthesis-design of biorefinery networks is strongly affected by the geopolitical and social situations that change readily. This has two main implications: (i) fast eval-

uations need to be performed in this dynamic context, which is addressed thought the development of a synthesis tool; and (ii) the impact of government incentives and subsidies as well as the impact on society of greener and more sustainable processes can be included in early stage assessments, since these factors can represent key contributors towards the establishment of biorefineries, given that economic profitability is not always comparable to traditional petroleum based processes.

Appendices

APPENDIX A

Review table

Table A.1. Review of works containing "biorefinery" and "superstructure optimization" in the period 2014-2016.

Study	Scope	Feedstocks, products, technologies, locations	Objective function and size	Main results	Models, problem type and solver
		2016			
Giuliano et al. [106]	Process flowsheet optimization. Sensitivity analysis.	F: Lignocellulose (hardwood). Rate 50 t/h (also 5 t/h, 150 t/h). P: Levulinic acid, succinic acid, ethanol. T: Thermochemical, biochemical.	OF: Economic (NPV, IRR). NF=1; NS=6; NI=33; NP=3	Most sensitive to economic scenario and plant scale. OF choice affects solution.	M: Shortcut. P: MILP (linearized MINLP). S: CPLEX
Martín et al. [107]	Optimal integrated topology, including technologies for thermal energy and methanol production.	F: Switchgrass, algae. P: Biodiesel (FAME). Rate 776 ML/y. T: Indirect gasification, steam reforming. For thermal energy and methanol: gasification and reforming.	OF: Economic (operating profit). NF=2; NI=20; NP=2	776 ML/year biofuels. 0 CO2 emissions, 1.27 kg CO2/kg meOH is captured.	M: Rigorous. P: NLP (decomposed MINLP). S: -
Yuan et al. [108]	Synthesis of integrated biorefinery.	F: Biomass. P: Liquid fuels and propylene. T: Gasification, pyrolysis, FT-synthesis, hydrotreating, upgrading.	OF: Min total cost (operating and capital). NF=1; NI=85; NP=4	-	M: Surrogate (ALAMO) and rigorous. P: Nonconvex MINLP. S: BARON (10%, 2-16 h)
Mountraki et al. [109]	Integrated treatment of biorefinery effluents. Generic bipartite graphs. Includes waste treatment. Central and distributed treatment. Grassroots and retrofit.	F: Wheat straw. P: Ethanol, ethylene, itaconic acid, xylitol, bio-PVC, PUs, PF Resins. T: Biochemical, thermochemical, etc.	OF: Min TAC. NF=1; NI=22; NF=7	3 examples: centralized vs decentralized, integrated solid and gas treatment, retrofitting existing installations.	M: Input-output models from regression. P: MINLP (bilinear terms). S: BARON (10%, 0.5-2.5 min)
Matthews et al. [110]	Process synthesis with fill heat, power and water integration. Parametric analysis.	F: Lignocellulose. P: Gasoline, diesel, jet-fuel (kerosene). T: Biological, thermochemical.	OF: Min costs (operating). NF=1; NI=125; NP=3	Bioconversion not economically competitive, except combined.	M: Input-output. P: MINLP. S: B&B Global Opt. (CPLEX, CONOPT)
Satchatippavarn et al. [111]	Design of urban biorefinery, integrating plastics and paper recycling.	F: urban waste (organic, plastics, paper). Rate 500,000 t/y. P: electricity, heat, compost, plastics, oil, recycled paper. T: AD + CHP, composting, incineration, mech. Recycling, pyrolysis, paper recycling. L: Bangkok.	OF: Max TP (total profit) NF=3; NI=6; NP=6	3 scenarios: recycling and tech-specific targets, supply of local demands, without incentives.	M: input-output. P: LP. S: - (GAMS).

Continued on next page

Table A.1 – continued from previous page

Study	Scope	Feedstocks, products, technologies, locations	Objective function and size	Main results	Models, problem type and solver
Giuliano et al. [112]	Optimal flowsheet from superstructure. Study of different seasonal biomass types during year.	F: eucalyptus, wheat straw, olive pruning. P: ethanol, succinic acid. T: biochemical, thermochemical.	OF: NPV, IRR. NF=3; NI=33; NP=3	Effect of composition of biomass feedstock.	M: Shortcut. P: MILP (linearized MINLP). S: CPLEX
		2015			
González-Delgado et al. [41]	Combined synthesis and analysis of topological pathways based on hierarchical and MP process synthesis.	F: microalgae. P: diesel-like fuel. T: transesterification, thermochemical, FT.	OF: max revenue. NF=1; NI=27; NP=1	-	M: Shortcut. P: - S: -
Rizwan et al. [40]	Superstructure optimization of microalgae production process including processing of microalgae residue.	F: microalgae (Chlorella vulgaris). P: biodiesel, glycerol, bio-oil, bio-ethanol, biogas. T: transesterification.	OF: yield, GOM. NF=1; NI=44; NP=5	2 scenarios: max of yield and max of GOM. Ranking of solutions (using integer cuts). Sensitivity analysis.	M: Shortcut. P: MINLP S: DICOPT
Cheali et al. [113]	Superstructure optimization under uncertainty with economic risk analysis.	F: corn stover, poplar wood. P: gasoline, diesel, ethylene, acetaldehydes, ethanol, acetone, butanol, blends, succinic acid. T: thermochemical, biochemical, ethanol upgrading.	OF: operating profit (EBITDA), sustainability single index ratio. NF=2; NI=92; NP=28	Bioethanol upgrading improves economics and sustainability of biorefinery. Multi-product biorefinery helps minimizing risk against price fluctuations.	M: Shortcut. P: MILP S: CPLEX
Cheali et al. [114]	Superstructure optimization of microalgae biorefinery.	F: race way algae. P: biodiesel, glycerol, gasoline, fertilizer, animal feed, biogas, bioethanol. T: transesterification.	OF: operating profit (EBITDA). NF=1; NI=31; NP=7	Deterministic solutions are different from under uncertainty. Cost of microalgae represents 90% of biodiesel production cost.	M: Shortcut. P: MILP S: -
Yue and You 2015					
		2014			

Continued on next page

Table A.1 – continued from previous page

Study	Scope	Feedstocks, products, technologies, locations	Objective function and size	Main results	Models, problem type and solver
Cheali et al. [115]	Process synthesis with uncertainty analysis for lignocellulosic biorefinery.	F: - P: - T: thermochemical, biochemical	OF: flowrate FT-products, FT-EBITDA, ethanol flowrate, ethanol-EBITDA NF=; NI=; NP=	Market price uncertainties reveal significant economic risks.	M: Shortcut. P: MINLP S: DICOPT
Kelloway et al. [116]	Systematic biorefinery synthesis to select products and technologies with highest economic potential or carbon efficiency. Optimal configuration, sensitivity analysis, Monte Carlo sampling, robustness analysis of selected configuration, MOO.	F: corn stover, wheat straw, barley straw, switchgrass. P: FT liquids, butane, alkene, succinic acid, acetic acid, xylitol, levulinic acid, formic acid. T: thermochemical, biochemical.	OF: NPV, carbon efficiency. NF=; NI=; NP=	Xylitol, levulinic acid and formic acid most favorable. Fuel prices play an important role. MOO shows the importance of multi-product biorefineries.	M: Detailed kinetics and shortcut. P: MINLP S: BARON
Gong et al. [117]	Superstructure of algal biorefinery processes for biological carbon sequestration and utilization.	F: CO2. P: diesel, power. T: cultivation, transesterification.	OF: profit per unit mass carbon dioxide sequestrated. NF=1; NI=34; NP=3	Minimum unit carbon sequestration cost of $1.48/ton CO2.	M: Linear conservation relationships. P: MINLP S: Branch-and-refine algorithm (CPLEX for MILP only required)
de Faria et al. [118]	Systematic synthesis and analysis applied to castor oil biorefinery.	F: castor oil. P: biodiesel, 12-hydroxystearic acid, hydrogenated castor oil, undecylenic acid, heptaldehyde, glycerol, 2-octanol, sebacic acid, ricinoleic acid. T: transesterification, thermochemical.	OF: profit. NF=1; NI=75; NP=10	Optimal product portfolio and flowsheet identified.	M: Shortcut. P: MILP S: CPLEX
Giuliano et al. [119]	Superstructure optimization of lignocellulose to levulinic acid, succinic acid and ethanol.	F: lignocellulose. P: levulinic acid, succinic acid, ethanol. T: biochemical.	OF: profit. NF= 1; NI=-; NP=3.	Optimal product distribution was found as well as flowsheet and operating conditions.	M: Shortcut. P: MILP (discretized MINLP) S: CPLEX

Continued on next page

Table A.1 – continued from previous page

Study	Scope	Feedstocks, products, technologies, locations	Objective function and size	Main results	Models, problem type and solver
Gong et al. [120]	Process superstructure of algal biorefinery for biological carbon sequestration and HC biofuel production.	F: CO2. P: diesel, power. T: cultivation, transesterification.	OF: profit per unit mass carbon dioxide sequestrated. NF=1; NI=34; NP=3	Minimum unit carbon sequestration cost of $1.48/ton CO2.	M: Linear conservation relationships. P: MINLP S: Branch-and-refine algorithm (CPLEX for MILP only required)
Zhang et al. [121]	Synthesis and design of hydrocarbon biorefinery under economic and environmental considerations, MOO via epsilon-constraint method.	F: biomass. P: diesel, gasoline. T: thermochemical.	OF: NPV, GWP. NF=1; NI=34; NP=3	Pareto curve indicates trade-off between NPV and GWP, two optimal designs identified.	M: Linear conservation relationships. P: MINLP S: BARON
Tsakalova et al. [122]	Screening of second generation biorefinery paths for energy and bio-based products	F: lignocellulosic biomass, wood residues, plant oils, MSW, animal fats/cooked oil. P: ethylene. T: transesterification, thermochemical.	OF: profit. NF=4; NI=36; NP=6		M: Shortcut. P: MILP S: -
Slegers et al. [123]	Combinatorial optimization approach for processing of microalgae.	F: algae solution. P: biodiesel. T: various.	OF: net energy ratio (NER: ratio biodiesel yield to energy consumption). NF=1; NI=18; NP=1	Investigate fixed and optimized process conditions. *Claim global optimum but do not mention solution strategy.	M: Shortcut. P: - S: -

APPENDIX B

Molasses example – data & results

B.1 Additional data

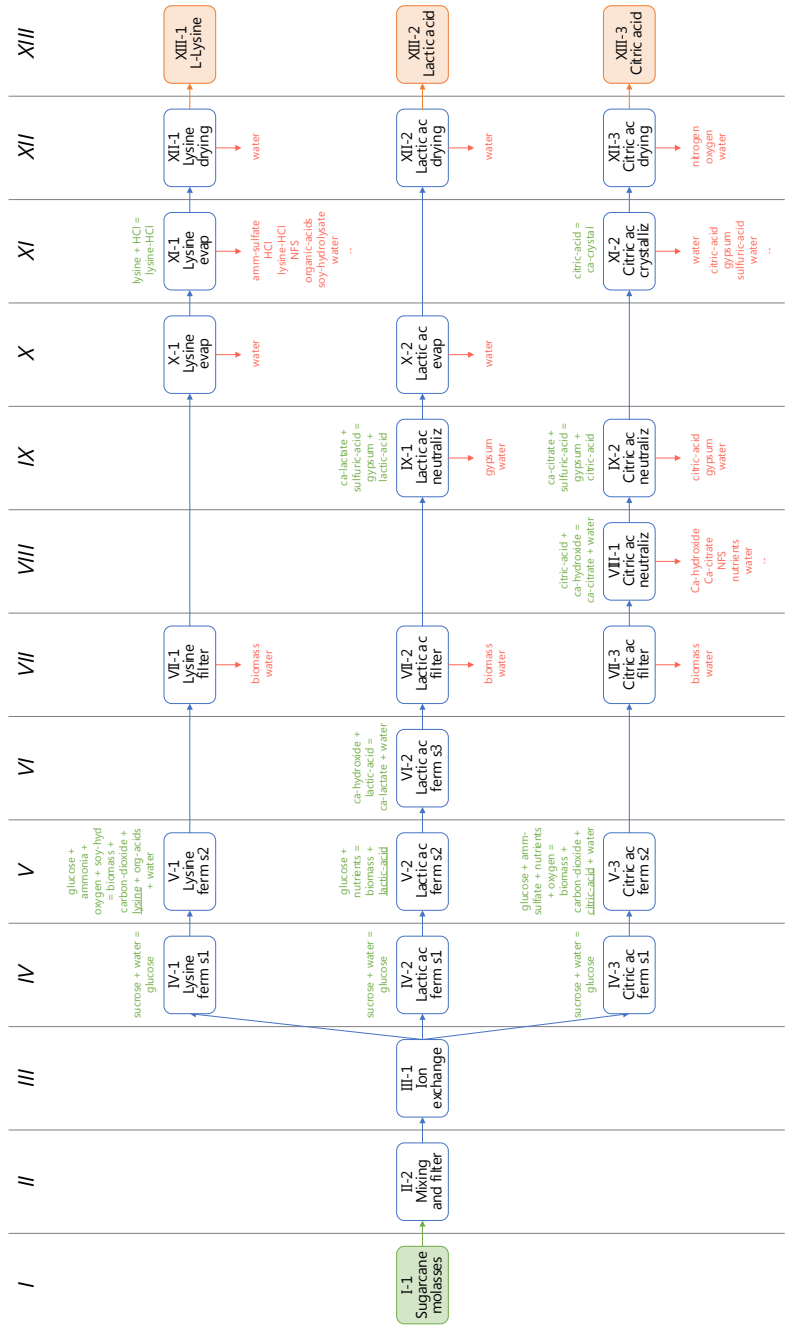


Figure B.1. Superstructure of alternatives for the sugarcane molasses example with reaction and waste data.

Table B.1. Molasses reaction stoichiometry data: $\gamma^{i,r}$.

	L-1	L-2	L-3	CA-1	CA-2	CA-3	CA-4	CA-5	LA-1	LA-2	LA-3	LA-4
sucrose	-1											
water	-1											
glucose	2											
glucose		-1										
ammonia		-1.12										
oxygen		-4.08										
soy-hydrolysate		-0.08										
biomass		1.24										
carbon-dioxide		3.16										
lysine		0.6										
organic-acids		0.48										
water		4.35										
lysine			-1									
HCl			-1									
lysine-HCl			1									
sucrose				-1								
water				-1								
glucose				2								
glucose					-1							
amm-sulfate					-0.02							
nutrients					-0.06							
oxygen					-1.75							
biomass					0.65							
carbon-dioxide					0.68							
citric-acid					0.8							
water					2.78							
citric-acid						-1						
ca-hydroxide						-1.5						
ca-citrate						0.5						
water						3						
ca-citrate							-1					
sulfuric-acid							-3					
gypsum							3					
citric-acid							2					
citric-acid								-1				
ca-crystal								1				
sucrose									-1			
water									-1			
glucose									2			
glucose										-1		
nutrients										-0.06		
biomass										0.49		
lactic-acid										2		
ca-hydroxide											-1	
lactic-acid											-2	
ca-lactate											1	
water											1	
ca-lactate												-1
sulfuric-acid												-1
gypsum												1
lactic-acid												2

B.2 Selected routes

B.3 Results scenario M-1

- Inlet flow rate (Table B.2)
- Added chemicals flow rate (Table B.3)
- Post-mixing flow rate (Table B.4)
- Post-reaction flow rate (Table B.5)
- Post-waste removal flow rate (Table B.6)
- Waste flow rate (Table B.7)
- Utility consumption rate (Table B.9)
- Total interval to interval flow rate (Table B.10)

Table B.2. Molasses results scenario M-1: $f_{IN}^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60
biomass				2.17	1810.40
ca-citrate					
ca-crystal					
citric-acid					17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67
oxygen				59879.58	
sucrose	14000.00	13971.90	13971.90	139.72	139.72
sulfuric-acid					
water	11600.00	71456.30	71456.30	91561.45	97212.32
	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate		22016.70	110.08		
ca-crystal				15878.74	15878.74
citric-acid	17405.05		16717.94		
glucose	205.48				
gypsum			896.55		
impurities					
NFS	7983.94				
nutrients	568.67				
oxygen					
sucrose	139.72				
sulfuric-acid			517.26		
water	93791.19	9219.88	131780.16	6858.78	80.68

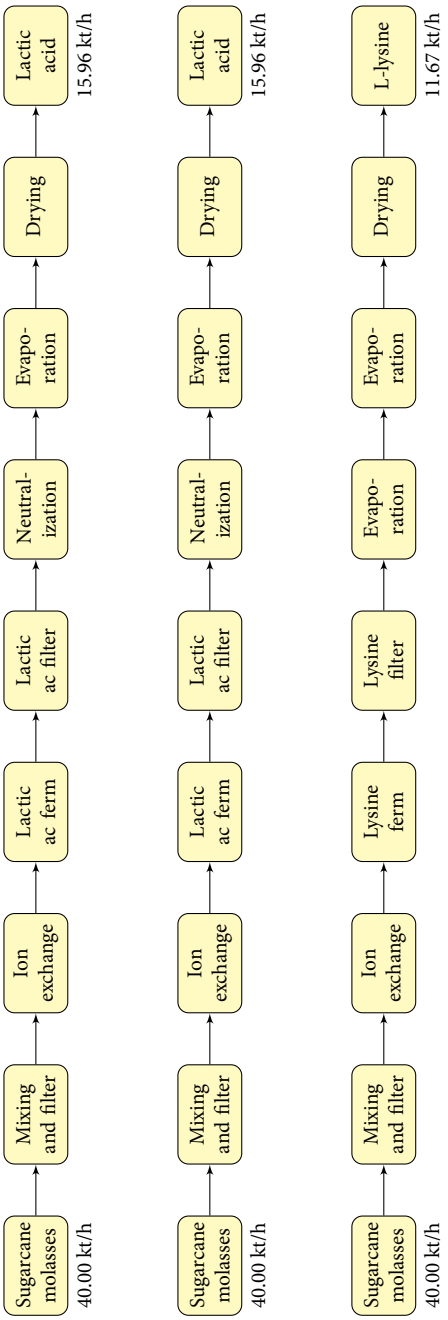


Figure B.2. Selected processing routes for each scenario of sugarcane molasses conversion to value-adedd chemicals.

Table B.3. Molasses results scenario M-1: $g_M^{i,k}$.

	II-1	III-1	IV-2	VII-2	VIII-1	IX-1	XI-2	XII-2
amm-sulfate			282.09					
biomass			2.17					
ca-hydroxide					10519.51			
nitrogen								5142.37
nutrients			1924.88					
oxygen			59879.58					1561.12
sodium-hydroxide		104.53						
sulfuric-acid						13448.69		
water	60000.00	5284.26	20833.33	7315.61	39967.12	129933.13	10315.49	

Table B.4. Molasses results scenario M-1: $f_M^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate			282.09	282.09	56.60
biomass			2.17	2.17	1810.40
ca-citrate					
ca-crystal					
ca-hydroxide					
citric-acid					17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nitrogen					
nutrients			1924.88	1924.88	568.67
oxygen			59879.58	59879.58	
sodium-hydroxide		104.53			
sucrose	14000.00	13971.90	13971.90	139.72	139.72
sulfuric-acid					
water	71600.00	76740.55	92289.63	91561.45	104527.93

	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate		22016.70	110.08		
ca-crystal				15878.74	15878.74
ca-hydroxide	10519.51				
citric-acid	17405.05		16717.94		
glucose	205.48				
gypsum			896.55		
impurities					
NFS	7983.94				
nitrogen				5142.37	
nutrients	568.67				
oxygen				1561.12	
sodium-hydroxide					
sucrose	139.72				
sulfuric-acid		13448.69	517.26		
water	133758.32	139153.01	142095.65	6858.78	80.68

Table B.5. Molasses results scenario M-1: $f_R^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate			282.09	56.60	56.60
biomass			2.17	1810.40	1810.40
ca-citrate					
ca-crystal					
ca-hydroxide					
carbon-dioxide				3390.70	
citric-acid				17405.05	17405.05
glucose	6000.00	5987.96	20548.36	205.48	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nitrogen					
nutrients			1924.88	568.67	568.67
oxygen			59879.58	53550.61	
sodium-hydroxide		104.53			
sucrose	14000.00	13971.90	139.72	139.72	139.72
sulfuric-acid					
water	71600.00	76740.55	91561.45	97212.32	104527.93
	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate	22466.02	110.08	110.08		
ca-crystal			16202.79	15878.74	15878.74
ca-hydroxide	500.93				
carbon-dioxide					
citric-acid	87.03	16886.81	334.36		
glucose	205.48				
gypsum		17931.03	896.55		
impurities					
NFS	7983.94				
nitrogen				5142.37	
nutrients	568.67				
oxygen				1561.12	
sodium-hydroxide					
sucrose	139.72				
sulfuric-acid		517.26	517.26		
water	138631.38	139153.01	142095.65	6858.78	80.68

Table B.6. Molasses results scenario M-1: $f_W^{i,k}$.

	I-1	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60	56.60
biomass				2.17	1810.40	
ca-citrate						
ca-crystal						
citric-acid					17405.05	17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48	205.48
gypsum						
impurities	400.00	40.00				
NFS	8000.00	7983.94	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67	568.67
oxygen				59879.58		
sucrose	14000.00	13971.90	13971.90	139.72	139.72	139.72
sulfuric-acid						
water	11600.00	71456.30	71456.30	91561.45	97212.32	93791.19

	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate					
biomass					
ca-citrate	22016.70	110.08			
ca-crystal			15878.74	15878.74	15878.74
citric-acid		16717.94			
glucose					
gypsum		896.55			
impurities					
NFS					
nutrients					
oxygen					
sucrose					
sulfuric-acid		517.26			
water	9219.88	131780.16	6858.78	80.68	80.68

Table B.7. Molasses results scenario M-1: $g_W^{i,k}$.

	II-1	III-1	V-2	VII-2	VIII-1	IX-1	XI-2	XII-2
amm-sulfate					56.60			
biomass				1810.40				
ca-citrate					449.32		110.08	
ca-crystal							324.06	
ca-hydroxide					500.93			
carbon-dioxide			3390.70					
citric-acid					87.03	168.87	334.36	
glucose	12.04				205.48			
gypsum						17034.48	896.55	
impurities	360.00	40.00						
NFS	16.06				7983.94			
nitrogen								5142.37
nutrients					568.67			
oxygen			53550.61					1561.12
sodium-hydroxide		104.53						
sucrose	28.10				139.72			
sulfuric-acid							517.26	
water	143.70	5284.26		10736.74	129411.50	7372.85	135236.87	6778.10

Table B.8. Molasses results scenario M-1: $f_{OUT,1}^{i,k}$.

	I-1	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60	56.60
biomass				2.17	1810.40	
ca-citrate						
ca-crystal						
citric-acid					17405.05	17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48	205.48
gypsum						
impurities	400.00	40.00				
NFS	8000.00	7983.94	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67	568.67
oxygen				59879.58		
sucrose	14000.00	13971.90	13971.90	139.72	139.72	139.72
sulfuric-acid						
water	11600.00	71456.30	71456.30	91561.45	97212.32	93791.19
	VIII-1	IX-1	XI-2	XII-2	XIII-2	
amm-sulfate						
biomass						
ca-citrate	22016.70	110.08				
ca-crystal			15878.74	15878.74	15878.74	
citric-acid		16717.94				
glucose						
gypsum		896.55				
impurities						
NFS						
nutrients						
oxygen						
sucrose						
sulfuric-acid		517.26				
water	9219.88	131780.16	6858.78	80.68	80.68	

Table B.9. Molasses results scenario M-1: $g_{UTIL}^{ut,k}$.

	IV-2	V-2	VIII-1	IX-1	XI-2	XII-2
power	11805.61	9566.57	29.80	46.85	80.98	368.11
steam				13.56	96.93	
chilled-water					920.98	
cooled-water	950.43	12120.28	2.58			

Table B.10. Molasses results scenario M-1: $F_{INT}^{k,kk}$.

	II-1	III-1	IV-2	V-2	VII-2
I-1	40000.00				
II-1		99440.10			
III-1			99400.10		
IV-2				182322.18	
V-2					125382.19
VII-2					
VIII-1					
IX-1					
XI-2					
XII-2					

	VIII-1	IX-1	XI-2	XII-2	XIII-2
I-1					
II-1					
III-1					
IV-2					
V-2					
VII-2	120150.66				
VIII-1		31236.58			
IX-1			150021.99		
XI-2				22737.51	
XII-2					15959.42

B.4 Results scenario M-2

- Inlet flow rate (Table B.11)
- Added chemicals flow rate (Table B.12)
- Post-mixing flow rate (Table B.13)
- Post-reaction flow rate (Table B.14)
- Post-waste removal flow rate (Table B.15)
- Waste flow rate (Table B.16)
- Utility consumption rate (Table B.18)
- Total interval-to-interval flow rate (Table B.19)

Table B.11. Molasses results scenario M-2: $f_{IN}^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60
biomass				2.17	1810.40
ca-citrate					
ca-crystal					
citric-acid					17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67
oxygen				59879.58	
sucrose	14000.00	13971.90	13971.90	139.72	139.72
sulfuric-acid					
water	11600.00	71456.30	71456.30	91561.45	97212.32
	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate		22016.70	110.08		
ca-crystal				15878.74	15878.74
citric-acid	17405.05		16717.94		
glucose	205.48				
gypsum			896.55		
impurities					
NFS	7983.94				
nutrients	568.67				
oxygen					
sucrose	139.72				
sulfuric-acid			517.26		
water	93791.19	9219.88	131780.16	6858.78	80.68

Table B.12. Molasses results scenario M-2: $g_M^{i,k}$.

	II-1	III-1	IV-2	VII-2	VIII-1	IX-1	XI-2	XII-2
amm-sulfate			282.09					
biomass			2.17					
ca-hydroxide					10519.51			
nitrogen								5142.37
nutrients			1924.88					
oxygen			59879.58					1561.12
sodium-hydroxide		104.53						
sulfuric-acid						13448.69		
water	60000.00	5284.26	20833.33	7315.61	39967.12	129933.13	10315.49	

Table B.13. Molasses results scenario M-2: $f_M^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate			282.09	282.09	56.60
biomass			2.17	2.17	1810.40
ca-citrate					
ca-crystal					
ca-hydroxide					
citric-acid					17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nitrogen					
nutrients			1924.88	1924.88	568.67
oxygen			59879.58	59879.58	
sodium-hydroxide		104.53			
sucrose	14000.00	13971.90	13971.90	139.72	139.72
sulfuric-acid					
water	71600.00	76740.55	92289.63	91561.45	104527.93

	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate		22016.70	110.08		
ca-crystal				15878.74	15878.74
ca-hydroxide	10519.51				
citric-acid	17405.05		16717.94		
glucose	205.48				
gypsum			896.55		
impurities					
NFS	7983.94				
nitrogen				5142.37	
nutrients	568.67				
oxygen				1561.12	
sodium-hydroxide					
sucrose	139.72				
sulfuric-acid		13448.69	517.26		
water	133758.32	139153.01	142095.65	6858.78	80.68

Table B.14. Molasses results scenario M-2: $f_R^{i,k}$.

	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate			282.09	56.60	56.60
biomass			2.17	1810.40	1810.40
ca-citrate					
ca-crystal					
ca-hydroxide					
carbon-dioxide				3390.70	
citric-acid				17405.05	17405.05
glucose	6000.00	5987.96	20548.36	205.48	205.48
gypsum					
impurities	400.00	40.00			
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
nitrogen					
nutrients			1924.88	568.67	568.67
oxygen			59879.58	53550.61	
sodium-hydroxide		104.53			
sucrose	14000.00	13971.90	139.72	139.72	139.72
sulfuric-acid					
water	71600.00	76740.55	91561.45	97212.32	104527.93
	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate	56.60				
biomass					
ca-citrate	22466.02	110.08	110.08		
ca-crystal			16202.79	15878.74	15878.74
ca-hydroxide	500.93				
carbon-dioxide					
citric-acid	87.03	16886.81	334.36		
glucose	205.48				
gypsum		17931.03	896.55		
impurities					
NFS	7983.94				
nitrogen				5142.37	
nutrients	568.67				
oxygen				1561.12	
sodium-hydroxide					
sucrose	139.72				
sulfuric-acid		517.26	517.26		
water	138631.38	139153.01	142095.65	6858.78	80.68

Table B.15. Molasses results scenario M-2: $f_W^{i,k}$.

	I-1	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60	56.60
biomass				2.17	1810.40	
ca-citrate						
ca-crystal						
citric-acid					17405.05	17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48	205.48
gypsum						
impurities	400.00	40.00				
NFS	8000.00	7983.94	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67	568.67
oxygen				59879.58		
sucrose	14000.00	13971.90	13971.90	139.72	139.72	139.72
sulfuric-acid						
water	11600.00	71456.30	71456.30	91561.45	97212.32	93791.19

	VIII-1	IX-1	XI-2	XII-2	XIII-2
amm-sulfate					
biomass					
ca-citrate	22016.70	110.08			
ca-crystal			15878.74	15878.74	15878.74
citric-acid		16717.94			
glucose					
gypsum		896.55			
impurities					
NFS					
nutrients					
oxygen					
sucrose					
sulfuric-acid		517.26			
water	9219.88	131780.16	6858.78	80.68	80.68

Table B.16. Molasses results scenario M-2: $g_W^{i,k}$.

	II-1	III-1	V-2	VII-2	VIII-1	IX-1	XI-2	XII-2
amm-sulfate					56.60			
biomass				1810.40				
ca-citrate					449.32		110.08	
ca-crystal							324.06	
ca-hydroxide					500.93			
carbon-dioxide			3390.70					
citric-acid					87.03	168.87	334.36	
glucose	12.04				205.48			
gypsum						17034.48	896.55	
impurities	360.00	40.00						
NFS	16.06				7983.94			
nitrogen								5142.37
nutrients					568.67			
oxygen			53550.61					1561.12
sodium-hydroxide		104.53						
sucrose	28.10				139.72			
sulfuric-acid							517.26	
water	143.70	5284.26		10736.74	129411.50	7372.85	135236.87	6778.10

Table B.17. Molasses results scenario M-2: $f^{i,k}_{OUT,1}$.

	I-1	II-1	III-1	IV-2	V-2	VII-2
amm-sulfate				282.09	56.60	56.60
biomass				2.17	1810.40	
ca-citrate						
ca-crystal						
citric-acid					17405.05	17405.05
glucose	6000.00	5987.96	5987.96	20548.36	205.48	205.48
gypsum						
impurities	400.00	40.00				
NFS	8000.00	7983.94	7983.94	7983.94	7983.94	7983.94
nutrients				1924.88	568.67	568.67
oxygen				59879.58		
sucrose	14000.00	13971.90	13971.90	139.72	139.72	139.72
sulfuric-acid						
water	11600.00	71456.30	71456.30	91561.45	97212.32	93791.19
	VIII-1	IX-1	XI-2	XII-2	XIII-2	
amm-sulfate						
biomass						
ca-citrate	22016.70	110.08				
ca-crystal			15878.74	15878.74	15878.74	
citric-acid		16717.94				
glucose						
gypsum		896.55				
impurities						
NFS						
nutrients						
oxygen						
sucrose						
sulfuric-acid		517.26				
water	9219.88	131780.16	6858.78	80.68	80.68	

Table B.18. Molasses results scenario M-2: $g^{ut,k}_{UTIL}$.

	IV-2	V-2	VIII-1	IX-1	XI-2	XII-2
power	11805.61	9566.57	29.80	46.85	80.98	368.11
steam				13.56	96.93	
chilled-water					920.98	
cooled-water	950.43	12120.28	2.58			

Table B.19. Molasses results scenario M-2: $F_{INT}^{k,kk}$.

	II-1	III-1	IV-2	V-2	VII-2
I-1	40000.00				
II-1		99440.10			
III-1			99400.10		
IV-2				182322.18	
V-2					125382.19
VII-2					
VIII-1					
IX-1					
XI-2					
XII-2					

	VIII-1	IX-1	XI-2	XII-2	XIII-2
I-1					
II-1					
III-1					
IV-2					
V-2					
VII-2	120150.66				
VIII-1		31236.58			
IX-1			150021.99		
XI-2				22737.51	
XII-2					15959.42

B.5 Results scenario M-3

- Inlet flow rate (Table B.20)
- Added chemicals flow rate (Table B.21)
- Post-mixing flow rate (Table B.22)
- Post-reaction flow rate (Table B.23)
- Post-waste removal flow rate (Table B.24)
- Waste flow rate (Table B.25)
- Utility consumption rate (Table B.27)
- Total interval-to-interval flow rate (Table B.28)

Table B.20. Molasses results scenario M-3: $f_{IN}^{i,k}$.

	II-1	III-1	IV-1	V-1	VII-1
amm-acetate				6.00	6.00
ammonia				4833.33	
amm-sulfate				1566.67	1566.67
biomass				3.00	3442.48
glucose	6000.00	5987.96	5987.96	20548.36	205.48
impurities	400.00	40.00			
KH2PO4				190.00	190.00
lysine					9826.63
lysine-HCl					
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
organic-acids					982.74
oxygen				59879.58	
soy-hydrolysate				3500.00	1927.84
sucrose	14000.00	13971.90	13971.90	139.72	139.72
water	11600.00	71456.30	71456.30	89394.78	98239.43

	X-1	XI-1	XII-1	XIII-1
amm-acetate	6.00	6.00		
ammonia				
amm-sulfate	1566.67	1566.67		
biomass				
glucose	205.48	205.48		
impurities				
KH2PO4	190.00	190.00		
lysine	9826.63	9826.63		
lysine-HCl			11670.99	11670.99
NFS	7983.94	7983.94		
organic-acids	982.74	982.74		
oxygen				
soy-hydrolysate	1927.84	1927.84		
sucrose	139.72	139.72		
water	91677.54	70681.87	38.18	1.01

Table B.21. Molasses results scenario M-3: $g_{M}^{i,k}$.

	II-1	III-1	IV-1	VII-1	XI-1	XII-1
amm-acetate			6.00			
ammonia			4833.33			
amm-sulfate			1566.67			
biomass			3.00			
HCl					2898.08	
KH2PO4			190.00			
oxygen			59879.58			
sodium-hydroxide		104.53				
soy-hydrolysate			3500.00			
water	60000.00	5284.26	18666.67	14433.79	45058.83	77.78

Table B.22. Molasses results scenario M-3: $f_M^{i,k}$.

	II-1	III-1	IV-1	V-1	VII-1
amm-acetate			6.00	6.00	6.00
ammonia			4833.33	4833.33	
amm-sulfate			1566.67	1566.67	1566.67
biomass			3.00	3.00	3442.48
glucose	6000.00	5987.96	5987.96	20548.36	205.48
HCl					
impurities	400.00	40.00			
KH2PO4			190.00	190.00	190.00
lysine					9826.63
lysine-HCl					
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
organic-acids					982.74
oxygen			59879.58	59879.58	
sodium-hydroxide		104.53			
soy-hydrolysate			3500.00	3500.00	1927.84
sucrose	14000.00	13971.90	13971.90	139.72	139.72
water	71600.00	76740.55	90122.96	89394.78	112673.22

	X-1	XI-1	XII-1	XIII-1	
amm-acetate	6.00	6.00			
ammonia					
amm-sulfate	1566.67	1566.67			
biomass					
glucose	205.48	205.48			
HCl		2898.08			
impurities					
KH2PO4	190.00	190.00			
lysine	9826.63	9826.63			
lysine-HCl			11670.99	11670.99	
NFS	7983.94	7983.94			
organic-acids	982.74	982.74			
oxygen					
sodium-hydroxide					
soy-hydrolysate	1927.84	1927.84			
sucrose	139.72	139.72			
water	91677.54	115740.70	115.96	1.01	

Table B.23. Molasses results scenario M-3: $f_R^{i,k}$.

	II-1	III-1	IV-1	V-1	VII-1
amm-acetate			6.00	6.00	6.00
ammonia			4833.33	2671.39	
amm-sulfate			1566.67	1566.67	1566.67
biomass			3.00	3442.48	3442.48
carbon-dioxide				15723.51	
glucose	6000.00	5987.96	20548.36	205.48	205.48
HCl					
impurities	400.00	40.00			
KH2PO4			190.00	190.00	190.00
lysine				9826.63	9826.63
lysine-HCl					
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
organic-acids				982.74	982.74
oxygen			59879.58	45138.63	
sodium-hydroxide		104.53			
soy-hydrolysate			3500.00	1927.84	1927.84
sucrose	14000.00	13971.90	139.72	139.72	139.72
water	71600.00	76740.55	89394.78	98239.43	112673.22

	X-1	XI-1	XII-1	XIII-1	
amm-acetate	6.00	6.00			
ammonia					
amm-sulfate	1566.67	1566.67			
biomass					
carbon-dioxide					
glucose	205.48	205.48			
HCl		496.24			
impurities					
KH2PO4	190.00	190.00			
lysine	9826.63	196.53			
lysine-HCl		12031.95	11670.99	11670.99	
NFS	7983.94	7983.94			
organic-acids	982.74	982.74			
oxygen					
sodium-hydroxide					
soy-hydrolysate	1927.84	1927.84			
sucrose	139.72	139.72			
water	91677.54	115740.70	115.96	1.01	

Table B.24. Molasses results scenario M-3: $f_W^{i,k}$.

	I-1	II-1	III-1	IV-1	V-1
amm-acetate				6.00	6.00
ammonia				4833.33	
amm-sulfate				1566.67	1566.67
biomass				3.00	3442.48
glucose	6000.00	5987.96	5987.96	20548.36	205.48
impurities	400.00	40.00			
KH2PO4				190.00	190.00
lysine					9826.63
lysine-HCl					
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
organic-acids					982.74
oxygen				59879.58	
soy-hydrolysate				3500.00	1927.84
sucrose	14000.00	13971.90	13971.90	139.72	139.72
water	11600.00	71456.30	71456.30	89394.78	98239.43
	VII-1	X-1	XI-1	XII-1	XIII-1
amm-acetate	6.00	6.00			
ammonia					
amm-sulfate	1566.67	1566.67			
biomass					
glucose	205.48	205.48			
impurities					
KH2PO4	190.00	190.00			
lysine	9826.63	9826.63			
lysine-HCl			11670.99	11670.99	11670.99
NFS	7983.94	7983.94			
organic-acids	982.74	982.74			
oxygen					
soy-hydrolysate	1927.84	1927.84			
sucrose	139.72	139.72			
water	91677.54	70681.87	38.18	1.01	1.01

Table B.25. Molasses results scenario M-3: $g_W^{i,k}$.

	II-1	III-1	V-1	VII-1	X-1	XI-1	XII-1
amm-acetate						6.00	
ammonia			2671.39				
amm-sulfate						1566.67	
biomass				3442.48			
carbon-dioxide			15723.51				
glucose	12.04					205.48	
HCl						496.24	
impurities	360.00	40.00					
KH2PO4						190.00	
lysine						196.53	
lysine-HCl						360.96	
NFS	16.06					7983.94	
organic-acids						982.74	
oxygen			45138.63				
sodium-hydroxide		104.53					
soy-hydrolysate						1927.84	
sucrose	28.10					139.72	
water	143.70	5284.26		20995.68	20995.68	115702.52	114.95

Table B.26. Molasses results scenario M-3: $f^{i,k}_{OUT,1}$.

	I-1	II-1	III-1	IV-1	V-1
amm-acetate				6.00	6.00
ammonia				4833.33	
amm-sulfate				1566.67	1566.67
biomass				3.00	3442.48
glucose	6000.00	5987.96	5987.96	20548.36	205.48
impurities	400.00	40.00			
KH2PO4				190.00	190.00
lysine					9826.63
lysine-HCl					
NFS	8000.00	7983.94	7983.94	7983.94	7983.94
organic-acids					982.74
oxygen				59879.58	
soy-hydrolysate				3500.00	1927.84
sucrose	14000.00	13971.90	13971.90	139.72	139.72
water	11600.00	71456.30	71456.30	89394.78	98239.43
	VII-1	X-1	XI-1	XII-1	XIII-1
amm-acetate	6.00	6.00			
ammonia					
amm-sulfate	1566.67	1566.67			
biomass					
glucose	205.48	205.48			
impurities					
KH2PO4	190.00	190.00			
lysine	9826.63	9826.63			
lysine-HCl			11670.99	11670.99	11670.99
NFS	7983.94	7983.94			
organic-acids	982.74	982.74			
oxygen					
soy-hydrolysate	1927.84	1927.84			
sucrose	139.72	139.72			
water	91677.54	70681.87	38.18	1.01	1.01

Table B.27. Molasses scenario M-3: $g^{ut,k}_{UTIL}$.

	IV-1	V-1	X-1	XI-1	XII-1
power	2157.82	3340.18			232.22
steam			93.54		0.03
chilled-water			37.61	26.26	
cooled-water	173.72	597.83			

Table B.28. Molasses scenario M-3: $F_{INT}^{k,kk}$.

	II-1	III-1	IV-1	V-1	VII-1
I-1	40000.00				
II-1		99440.10			
III-1			99400.10		
IV-1				188045.38	
V-1					124510.94
VII-1					
X-1					
XI-1					
XII-1					

	X-1	XI-1	XII-1	XIII-1
I-1				
II-1				
III-1				
IV-1				
V-1				
VII-1	114506.57			
X-1		93510.89		
XI-1			11709.16	
XII-1				11672.00

Glossary

B

Biorefinery A biorefinery is defined in this thesis as a process or a network of processes that use biomass-derived feedstock to produce a series of products ranging from base and specialty chemicals to fuels and energy.

C

Connection Connections in the PSIN represent possible topological solutions of the synthesis problem. These connections can represent physical material or energy flow superstructure connections or conceptual location selection connections.

Superstructure connection Superstructure connections determine feasible combinations among processing intervals. Connections in the superstructure are directed, representing the flow direction and can connect elements of different type including feedstocks, technologies and products. Superstructure connections can be primary or secondary. Primary connections connect a primary outlet from the product separation task of an interval to other intervals. Secondary connections connect the secondary outlet from the same task to other intervals.

Location connection Location connections differ from the aforementioned in that they do not represent physical connections (therefore having neither flow of material nor of energy). Instead, they represent feasible relations between intervals and geographical areas, that is they connect intervals to locations to which they can be assigned.

I

Interval See *processing interval*

P

Processing interval A processing interval is defined as a set of tasks that represent an alternative to perform a given processing step. An interval might represent a set of unit operations, a single unit operation, or a part of a unit operation (i.e. multiple intervals can be used to represent a single unit) or phenomena. Processing intervals are represented as boxes (nodes) in the superstructure.

Processing section A processing section is defined as a group of processing intervals that are to be allocated in a single geographical location. The number of processing sections in a problem is defined by the user as well as the processing intervals that belong to each of them. An interval can only belong to one processing section and all intervals need to be placed in a section to solve location-dependent problems. Processing sections are represented as larger boxes containing sets of processing intervals.

Processing step A processing step represents a transformation step in the process, which can generally be realized through various alternatives. Processing steps are represented as columns in the superstructure. Commonly, the first step in a processing network is feedstock/source, followed by a sequence of processing steps and ending with a product step as last.

Processing task A processing task is defined as an elementary transformation that takes place in a process or unit operation incurring a physical, chemical or mechanical transformation on the stream(s) being processed. A set of elemental processing tasks are considered in this work, based on those defined by Zondervan et al. [39] and adopted by Quaglia et al. [20].

- *Mixing of chemicals* The mixing of chemicals refers to additional raw materials (other than the main biomass-derived feedstock), solvents, catalysts and other non-utility inlets added into the system.
- *Reaction* The reaction task allows the representation of tasks involving chemical transformations.
- *Waste removal* The waste removal task represents the separation of compounds that are removed from the system.
- *Product separation* The product separation task represents the separation of compounds that are to be further processed.
- *Utilities consumption* The consumption of utilities by the system for heating, cooling, power or other is represented by this task. The utilities added here are not mixed with the process stream.

R

Raw material See *feedstock*

S

Section See *processing section*

Step See *processing step*

T

Task See *processing task*

Nomenclature

Sets

c, cc	Integer cut
i, ii	Component
j, jj	Piecewise linearization
k, kk	Processing interval
l, ll	Geographical location
r	Reaction
st	Processing step
t, tt	Processing section (for allocation)
ut	Utility

Subsets

ADD_i	Chemical added (subset of component)
$PROC_k$	Processing interval (subset of interval)
$PROD_k$	Product interval (subset of interval)
RAW_k	Raw material/feedstock interval (subset of interval)
$REAC_i$	Key reactant (subset of component)

Parameters

$A^{k,l} \vert k \in RAW_k$	Availability of raw material k in location l	t/y
$Capacity$	Capacity	t/y
$D^{k,l} \vert k \in PROD_k$	Demand of product k in location l	t/y
M	Large value for big-M constraints	-
MW^i	Molar weight of component i	g/mol
MW^{reac}	Molar weight of key reactant $reac$	g/mol

$P^{i,l}_{CHEM} \vert i \in ADD_i$	Price of chemical i in location l	\$/t
$P^{k,l}_{PROD} \vert k \in PROD_k$	Price of product k in location l	\$/t
$P^{k,l}_{RAW} \vert k \in RAW_k$	Price of raw material k in location l	\$/t
P_{TRANS}	Price of transportation	\$/t/km
$P^{ut,l}_{UTIL}$	Price of utility ut in location l	kWh/t
P_{WASTE}	Price of waste handling	\$/t
α^k	Nonlinear capital cost function parameter	-
$\alpha^{j,k}_{LIN}$	Piecewise linearized capital cost function parameter	-
β^k	Nonlinear capital cost function parameter	-
$\beta^{j,k}_{LIN}$	Piecewise linearized capital cost function parameter	-
$\gamma^{i,r}$	Stoichiometric coefficient of component i in reaction r	-
$\delta^{i,k}$	Waste removal fraction of component i in interval k	-
$\zeta^{k,kk}$	Superstructure connection between interval k and interval kk	-
$\zeta^{k,kk}_P$	Primary connection between interval k and interval kk	-
$\zeta^{k,kk}_S$	Secondary connection between interval k and interval kk	-
$\eta^{l,ll}$	Distance from location l to location ll	km
$\theta^{reac,r,k}$	Conversion of reactant $reac$ in reaction r in interval k	-
$\vartheta^{k,l} \vert k \in RAW_k$	Moisture content of raw material k in location l	
$\lambda^{ut,k}_1$	Ratio of utility ut consumption in first utility consumption point of interval k	-
$\lambda^{ut,k}_2$	Ratio of utility ut consumption in second utility consumption point of interval k	-
$\lambda^{ut,k}_3$	Ratio of utility ut consumption in third utility consumption point of interval k	-
$\mu^{i,ii,k}$	Ratio of added chemical i in interval k with respect to reference component i	-
$\xi^{t,tt}$	Connection between section t and tt	-
$\sigma^{i,k}$	Separation recovery of component i in interval k	-
τ	Project lifetime	y
$\upsilon^{k,st}$	Allocation of interval k in step st	-
$\phi^{i,k,l} \vert k \in RAW_k$	Composition of component i in raw material k in location l	-
$\psi^{k,t}$	Allocation of interval k in processing section t	-
Γ^j		

Continuous variables

z	Objective function	\$/y
z_{EBIT}	Objective function (EBIT)	\$/y
z_{GOI}	Objective function (GOI)	\$/y

Positive continuous variables

C_{CAP}	Capital cost	\$
C_{CHEM}	Added chemicals cost	\$/y
c_{EQ}^{k}	Equipment cost of interval k	\$/y
C_{RAW}	Raw material cost	\$/y
C_{TRANS}	Transportation cost	\$/y
C_{UTIL}	Utility cost	\$/y
C_{WASTE}	Waste handling cost	\$/y
$f_1^{i,k,kk}$	Primary outlet flow rate of component i flowing from interval k to interval kk	t/y
$f_2^{i,k,kk}$	Secondary outlet flow rate of component i flowing from interval k to interval kk	t/y
F_{CAP}^{k}	Total flowrate in interval k for capital cost calculation	t/y
$F_D^{j,k}$		t/y
$f^{i,k,kk}$	Flow rate of component i flowing from interval k to interval kk	t/y
$f_{IN}^{i,k}$	Inlet flow rate of component i in interval k	t/y
$F_{INT}^{k,kk}$	Total flow rate between interval k and interval kk	t/y
$F_{LOC}^{l,ll}$	Total flow rate from location l to location ll	t/y
$f_M^{i,k}$	Post-mixing flow rate of component i in interval k	t/y
$f_M^{reac,k}$	Post-mixing flow rate of key reactant *reac* in interval k	t/y
$f_{OUT,1}^{i,k}$	Main separation product flow rate of component i in interval k	t/y
$f_{OUT,2}^{i,k}$	Secondary separation product flow rate of component i in interval k	t/y
$F_{PROD}^{k,l} \vert k \in PROD_k$	Total flow rate of product k to location l	t/y
$F_{RAW}^{k,l} \vert k \in RAW_k$	Total flow rate of raw material k to location l	t/y
$f_R^{i,k}$	Post-reaction flow rate of component i in interval k	t/y

$F_{SEC}^{t,tt}$	Total flow rate between section t and section tt	t/y
$f_{W,LOC}^{i,k,l}$	Flow rate of component i after waste separation task of interval k in location l	t/y
$f_{W}^{i,k}$	Post-waste separation flow rate of component i in interval k	t/y
$g_{M,LOC}^{i,k,l}$	Flow rate of added chemical i into interval k in location l	t/y
$g_{M}^{i,k}$	Flow rate of added chemical i into interval k	t/y
$g_{UTIL,LOC}^{ut,k,l}$	Consumption of utility ut in interval k in location l	kWh/y
$g_{UTIL}^{ut,k}$	Consumption of utility ut in interval k	kWh/y
$g_{W,LOC}^{i,k,l}$	Waste flow rate of component i in interval k	t/y
$g_{W}^{i,k}$	Waste flow rate of component i in interval k	t/y
$h^{t,tt,l,ll}$	Flow rate from section t in location l to section tt in location ll	t/y
N_{PROC}	Number of processing intervals (excludes feedstock and product)	-
N_{PROD}	Number of product intervals	-
N_{RAW}	Number of raw material/feedstock intervals	-
N_{SEC}	Number of processing sections (for allocation)	-
S_{PROD}	Product sales	$/y

Binary variables

$TR^{t,tt,l,ll}$	Transportation takes place between sections t and tt from location l to location ll	-
$w^{i,k}$	Selection of flowrate piece j in interval k	-
$x^{t,l}$	Allocation of section t to location l	-
y^{k}	Selection of interval k	-

Abbreviations

BR	Brazil
BSDB	Biorefinery synthesis database
CA	Canada
CAMD	Computer-aided molecular design
CARPS	Computer-aided reaction path synthesis
CN	China
CR	Cassava rhizome
CS	Corn stover
EBIT	Earning before interests and tax
ePSIN	Extended Processing Step-Interval Network
FAME	Fatty acid methyl ester
GAMS	Generic Algebraic Modeling System
GIS	Geographic information systems
GOI	Gross operating income
GOM	Gross operating margin
GWP	Global warming potential
HWC	Hardwood chips
IN	India
k	kilo (10^3)
LP	Linear programming
M	milion (10^6)
MILP	Mixed-integer linear programming
MINLP	Mixed-integer nonlinear programming
MIP	Mixed-integer (linear) programming
MX	Mexico
NLP	Nonlinear programming

PSE	Process Systems Engineering
NDV	Number of discrete variables
NEQ	Number of equations
NF	Number of raw materials/feedstocks
NI	Number of processing intervals (excluding feedstocks and products)
NL	Number of locations
NP	Number of products
NPV	Net present value
NV	Number of variables
PI	Process intensification
PSIN	Processing Step-Interval Network
SB	Sugarcane bagasse
SEN	State-Equipment Network
SG	Switch grass
SSCF	Simultaneous saccharification and fermentation
STN	State-Task Network
t	metric ton
TAC	Total annualized cost
TH	Thailand
US	United States of America
WC	Wood chips
WS	Wheat straw
y	year

Bibliography

[1] Nishida, N., Stephanopoulos, G., and Westerberg, A. W. A review of process synthesis. *AIChE Journal* **27** (1981), 321–351.

[2] Babi, D. K., Holtbruegge, J., Lutze, P., Gorak, A., Woodley, J. M., and Gani, R. Sustainable process synthesis–intensification. *Computers & Chemical Engineering* **81** (2015), 218–244.

[3] Fornari, T., Rotstein, E., and Stephanopoulos, G. Studies on the synthesis of chemical reaction paths—II. reaction schemes with two degrees of freedom. *Chemical Engineering Science* **44** (1989), 1569–1579.

[4] Barnicki, S. D. and Siirola, J. J. Process synthesis prospective. *Computers & Chemical Engineering* **28** (2004), 441–446.

[5] Li, X. and Kraslawski, A. Conceptual process synthesis: past and current trends. *Chemical Engineering and Processing: Process Intensification* **43** (2004), 583–594.

[6] Westerberg, A. W. A retrospective on design and process synthesis. *Computers and Chemical Engineering* **28** (2004), 447–458.

[7] Cremaschi, S. A perspective on process synthesis: Challenges and prospects. *Computers & Chemical Engineering* **81** (2015), 130–137.

[8] Chen, Q. and Grossmann, I. Recent Developments and Challenges in Optimization-Based Process Synthesis. *Annual Review of Chemical and Biomolecular Engineering* **8** (2017), 249–283.

[9] Douglas, J. M. A hierarchical decision procedure for process synthesis. *AIChE Journal* **31** (1985), 353–362.

[10] Singh, R., Gernaey, K. V., and Gani, R. An ontological knowledge-based system for the selection of process monitoring and analysis tools. *Computers & Chemical Engineering* **34** (2010), 1137–1154.

[11] King, C. J., Gantz, D. W., and Barnés, F. J. Systematic Evolutionary Process Systhesis. *Industrial & Engineering Chemistry Process Design and Development* **11** (1972), 271–283.

[12] McGalliard, R. L. and Westerberg, A. W. Structural sensitivity analysis in design synthesis. *The Chemical Engineering Journal* **4** (1972), 127–138.

[13] Stephanopoulos, G. and Westerberg, A. W. Studies in process synthesis—II: Evolutionary synthesis of optimal process flowsheets. *Chemical Engineering Science* **31** (1976), 195–204.

[14] Papoulias, S. a. and Grossmann, I. E. A structural optimization approach in process synthesis—I. *Computers & Chemical Engineering* **7** (1983), 695–706.

[15] Papalexandri, K. P. and Pistikopoulos, E. N. A process synthesis modelling framework based on mass/heat transfer module hyperstructure. *Computers & Chemical Engineering* **19** (1995), 71–76.

[16] D'Anterroches, L. Process Flowsheet Generation & Design Through a Group Contribution Approach. PhD thesis. Technical University of Denmark, 2005.

[17] Tula, A. K., Gani, R., and Eden, M. R. New Method and Software for Computer-Aided Flowsheet Design and Analysis. *Computer Aided Chemical Engineering*. Elsevier, 2017, 649–654.

[18] Yeomans, H. and Grossmann, I. E. A systematic modeling framework of superstructure optimization in process synthesis. *Computers & Chemical Engineering* **23** (1999), 709–731.

[19] Baliban, R. C., Elia, J. A., and Floudas, C. A. Optimization framework for the simultaneous process synthesis, heat and power integration of a thermochemical hybrid biomass, coal, and natural gas facility. *Computers & Chemical Engineering* **35** (2011), 1647–1690.

[20] Quaglia, A., Sarup, B., Sin, G., and Gani, R. Integrated business and engineering framework for synthesis and design of enterprise-wide processing networks. *Computers and Chemical Engineering* **38** (2012), 213–223.

[21] Sahinidis, N. V. BARON: A general purpose global optimization software package. *Journal of Global Optimization* **8** (1996), 201–205.

[22] Misener, R. and Floudas, C. A. Global optimization of mixed-integer quadratically-constrained quadratic programs (MIQCQP) through piecewise-linear and edge-concave relaxations. *Mathematical Programming* **136** (2012), 155–182.

[23] Grossmann, I. E., Caballero, J. A., and Yeomans, H. Mathematical programming approaches to the synthesis of chemical process systems. *Korean Journal of Chemical Engineering* **16** (1999), 407–426.

[24] Aspen Technology, Inc. AspenPlus. aspentech.com/.

[25] Schneider Electric. Simsci Pro/II. software.schneider-electric.com.

[26] Intelligen, Inc. SuperPro Designer. intelligen.com.

[27] Saengwirun, P. ECON: A software for cost calculation and economic analysis. MA thesis. The Petroleum and Petrochemical College, Chulalongkorn University, 2011.

[28] Piyarak, S. Development of software for Life Cycle Assessment. MA thesis. The Petroleum and Petrochemical College, Chulalongkorn University, 2012.

[29] Carvalho, A., Matos, H. A., and Gani, R. SustainPro—A tool for systematic process analysis, generation and evaluation of sustainable design alternatives. *Computers & Chemical Engineering* **50** (2013), 8–27.

[30] Gani, R. ICAS documentation. CAPEC, Technical University of Denmark. 2002.

[31] Schembecker, G., Simmrock, K. H., and Wolff, A. Synthesis of chemical process flowsheets by means of cooperating knowledge integrating systems. *Institution of Chemical Engineers Symposium Series* (1994).

[32] Schembecker, G. and Simmrock, K. H. Heuristic-numeric process synthesis with PROSYN. *AIChE symposium series* **92** (1996).

[33] Center, P. D. PROSYN. process-design-center.com.

[34] Ropotar, M. and Kravanja, Z. Implementation of efficient logic-based techniques in the MINLP process synthesizer MIPSYN. *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*. Elsevier, 2006, 233–238.

[35] Kravanja, Z. and Grossmann, I. New developments and capabilities in prosyn—An automated topology and parameter process synthesizer. *Computers and Chemical Engineering* **18** (1994), 1097–1114.

[36] Kravanja, Z. and Grossmann, I. PROSYN – an MINLP process synthesizer. *Computers & Chemical Engineering* **14** (1990), 1363–1378.

[37] Gassner, M. and Maréchal, F. Methodology for the optimal thermo-economic, multi-objective design of thermochemical fuel production from biomass. *Computers & Chemical Engineering* **33** (2009), 769–781.

[38] Martín, M. and Grossmann, I. E. Superstructure optimization of lignocellulosic bioethanol plants. *Computer Aided Chemical Engineering* **28** (2010), 943–948.

[39] Zondervan, E., Nawaz, M., Haan, A. B. de, Woodley, J. M., and Gani, R. Optimal design of a multi-product biorefinery system. *Computers & Chemical Engineering* **35** (2011), 1752–1766.

[40] Rizwan, M., Lee, J. H., and Gani, R. Optimal design of microalgae-based biorefinery: Economics, opportunities and challenges. *Applied Energy* **150** (2015), 69–79.

[41] González-Delgado, Á.-D., Kafarov, V., and El-Halwagi, M. Development of a topology of microalgae-based biorefinery: process synthesis and optimization using a combined forward–backward screening and superstructure approach. *Clean Technologies and Environmental Policy* **17** (2015), 2213–2228.

[42] Corbetta, M., Grossmann, I. E., and Manenti, F. Process simulator-based optimization of biorefinery downstream processes under the Generalized Disjunctive Programming framework. *Computers & Chemical Engineering* **88** (2016), 73–85.

[43] Martín, M. and Grossmann, I. E. BIOpt: A library of models for optimization of biofuel production processes. *Computer Aided Chemical Engineering* **30** (2012), 17–20.

[44] Iakovou, E., Karagiannidis, A., Vlachos, D., Toka, A., and Malamakis, A. Waste biomass-to-energy supply chain management: A critical synthesis. *Waste Management* **30** (2010), 1860–1870.

[45] Nikolopoulou, A. and Ierapetritou, M. G. Optimal design of sustainable chemical processes and supply chains: A review. *Computers & Chemical Engineering* **44** (2012), 94–103.

[46] Sharma, B., Ingalls, R., Jones, C., and Khanchi, A. Biomass supply chain design and analysis: Basis, overview, modeling, challenges, and future. *Renewable and Sustainable Energy Reviews* **24** (2013), 608–627.

[47] Elia, J. A. and Floudas, C. A. Energy Supply Chain Optimization of Hybrid Feedstock Processes: A Review. *Annual Review of Chemical and Biomolecular Engineering* **5** (2014), 147–179.

[48] Garcia, D. J. and You, F. Supply chain design and optimization: Challenges and opportunities. *Computers & Chemical Engineering* **81** (2015), 153–170.

[49] Rathore, A. S., Chopda, V. R., and Gomes, J. Knowledge management in a waste based biorefinery in the QbD paradigm. *Bioresource Technology* **215** (2016), 63–75.

[50] Holder, C. T., Cleland, J. C., LeDuc, S. D., Andereck, Z., Hogan, C., and Martin, K. M. Generating a geospatial database of U.S. regional feedstock production for use in evaluating the environmental footprint of biofuels. *Journal of the Air & Waste Management Association* **66** (2016), 356–365.

[51] Black, M., Sadhukhan, J., Day, K., Drage, G., and Murphy, R. Developing database criteria for the assessment of biomass supply chains for biorefinery development. *Chemical Engineering Research and Design* **107** (2016), 253–262.

[52] Balakshin, M. Y. and Capanema, E. A. Comprehensive structural analysis of biorefinery lignins with a quantitative 13C NMR approach. *RSC Advances* **5** (2015), 87187–87199.

[53] Barla, F., Lykokanellos, F., and Kokossis, A. C. Discovering Valorisation Paths in Waste Biorefineries using an Ontology Engineering Approach. *Computer Aided Chemical Engineering*. Elsevier, 2016, 2079–2084.

[54] Trokanas, N., Bussemaker, M., Velliou, E., Tokos, H., and Cecelja, F. BiOnto: An ontology for biomass and biorefining technologies. *Computer Aided Chemical Engineering* **37** (2015), 959–964.

[55] Siougkrou, E. and Kokossis, A. Development of semantically-enabled community hubs in biorefineries and biorenewables. *Computer Aided Chemical Engineering*. Elsevier, 2016, 2013–2018.

[56] Energy Research Center of the Netherlands (ECN) – Biomass. *BIODAT – Biomass database*. `biodat.eu`. (Online, accessed Sept 2016).

[57] C. Haslinger, K. R. an and M. Herger, H. H. *BIOBIB – A Database for biofuels*. `vt.tuwien.ac.at//biobib/EN/`. Institute of Chemical Engineering, Fuel and Environmental Technology, University of Technology Vienna. (Online, accessed Sept 2016).

[58] Biomass Energy Resource Center (BERC), Vermont Energy Investment Corporation (VEIC). *BERC Community-Scale Database*. `biomasscenter.org/database`. (Online, accessed Sept 2016).

[59] Ecostrat, University of Tennessee. *Wood2Energy*. `wood2energy.org/`. (Online, accessed Sept 2016).

[60] Energy research Center of the Netherlands (ECN), Biomass & Energy Efficiency. *Phyllis 2 – Database for biomass and waste*. `ecn.nl/phyllis2/`. (Online, accessed Sept 2016).

[61] Bioenergy Knowledge Discovery Framework (KDF), U.S. Department of Energy. *Bioenergy KDF Library*. `bioenergykdf.net/library`. (Online, accessed Sept 2016).

[62] SAHYOG: Strengthening Networking on Biomass Research and Biowaste Conversion – Biotechnology for Europe India Integration. *SAHYOG Biomass Database*. `sahyog-biomass-database.eu/`. (Online, accessed Sept 2016).

[63] U.S. Department of Energy, Idaho National Laboratory. *Bioenergy Feedstock Library*. `bioenergylibrary.inl.gov`. (Online, accessed Sept 2016).

[64] Bertran, M.-O., Frauzem, R., Sanchez-Arcilla, A.-S., Zhang, L., Woodley, J. M., and Gani, R. A generic methodology for processing route synthesis and design based on superstructure optimization. *Computers & Chemical Engineering* (2017).

[65] Duran, M. and Grossmann, I. A mixed-integer nonlinear programming algorithm for process systems synthesis. *AIChE Journal* **32** (1986), 592–606.

[66] Floudas, C. A. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press, 1995. 480.

[67] Perry, R. H. and Green, D. W. *Perry's Chemical Engineers' Handbook, 7th Edition*. McGraw-Hill, 1999.

[68] Lutze, P., Roman-Martinez, A., Woodley, J. M., and Gani, R. A systematic synthesis and design methodology to achieve process intensification in (bio) chemical processes. *Computers and Chemical Engineering* **36** (2012), 189–207.

[69] United Nations Environment Programme. *GEO-5 Global Environmental Outlook – Environment for the Future We Want*. Tech. rep. Nairobi, Kenya: United Nations Environment Programme, 2012.

[70] Grossmann, I. E. and Daichendt, M. M. New trends in optimization-based approaches to process synthesis. *Computers & Chemical Engineering* **20** (1996), 665–683.

[71] Quaglia, A. An Integrated Business and Engineering Framework for Synthesis and Design of Processing Networks. PhD thesis. Technical Univeristy of Denmark, 2013.

[72] GAMS Development Corporation. General Algebraic Modeling System (GAMS) Release 24.7.4. Washington, DC, USA. 2016.

[73] Nemhauser, G. and Wolsey, L. *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., 1988.

[74] Bergamini, M. L., Aguirre, P., and Grossmann, I. Logic-based outer approximation for globally optimal synthesis of process networks. *Computers & Chemical Engineering* **29** (2005), 1914–1933.

[75] Fazlollahi, S., Mandel, P., Becker, G., and Maréchal, F. Methods for multi-objective investment and operating optimization of complex energy systems. *Energy* **45** (2012), 12–22.

[76] Land, A. H. and Doig, A. G. An Automatic Method of Solving Discrete Programming Problems. *Econometrica* **28** (1960), 497.

[77] Kocis, G. R. and Grossmann, I. E. Relaxation strategy for the structural optimization of process flow sheets. *Industrial & Engineering Chemistry Research* **26** (1987), 1869–1880.

[78] Marrero, J. *ProPred Manual - CAPEC internal report*. Technical University of Denmark. Lyngby, Denmark, 2002.

[79] Villadsen, J., Nielsen, J., and Lidén, G. *Bioreaction Engineering Principles*. Springer US, 2011.

[80] Bek-Pedersen, E. and Gani, R. Design and synthesis of distillation systems using a driving-force-based approach. *Chemical Engineering and Processing: Process Intensification* **43** (2004), 251–262.

[81] Duran, M. A. and Grossmann, I. E. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming* **36** (1986), 307–339.

[82] Ahmed, S., Tawarmalani, M., and Sahinidis, N. V. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming* **100** (2004), 355–377.

[83] Voivontas, D., Assimacopoulos, D., and Koukios, E. Assessment of biomass potential for power production: a GIS based method. *Biomass and Bioenergy* **20** (2001), 101–112.

[84] Zhang, F., Johnson, D., Johnson, M., Watkins, D., Froese, R., and Wang, J. Decision support system integrating GIS with simulation and optimisation for a biofuel supply chain. *Renewable Energy* **85** (2016), 740–748.

[85] Cignitti, S. Computer-aided reaction path synthesis. MA thesis. Technical University of Denmark, 2014.

[86] Bertran, M.-O., Frauzem, R., Zhang, L., and Gani, R. A generic methodology for superstructure optimization of different processing networks. *Computer Aided Process Engineering* **38** (2016), 685–690.

[87] Quaglia, A., Pennati, A., Bogataj, M., Kravanja, Z., Sin, G., and Gani, R. Industrial process water treatment and reuse: A framework for synthesis and design. *Industrial & Engineering Chemistry Research* **53** (2014), 5160–5171.

[88] Anaya-Reza, O. and Lopez-Arenas, T. Comprehensive assessment of the L-lysine production process from fermentation of sugarcane molasses. *Bioprocess and Biosystems Engineering* **40** (2017), 1033–1048.

[89] Zhang, Y., Dubé, M. A., McLean, D. D., and Kates, M. Biodiesel production from waste cooking oil: 1. Process design and technological assessment. *Bioresource Technology* **89** (2003), 1–16.

[90] Rizwan, M., Zaman, M., Lee, J. H., and Gani, R. Optimal processing pathway selection for microalgae-based biorefinery under uncertainty. *Computers & Chemical Engineering* **82** (2015), 362–373.

[91] Rizwan, M., Lee, J. H., and Gani, R. Optimal processing pathway for the production of biodiesel from microalgal biomass: A superstructure based approach. *Computers & Chemical Engineering* **58** (2013), 305–314.

[92] Mansouri, S., Ismail, M., Babi, D., Simasatitkul, L., Huusom, J., and Gani, R. Systematic Sustainable Process Design and Analysis of Biodiesel Processes. *Processes* **1** (2013), 167–202.

[93] Simasatitkul, L., Gani, R., and Arpornwichanop, A. Optimal Design of Biodiesel Production Process from Waste Cooking Palm Oil. *Procedia Engineering* **42** (2012), 1292–1301.

[94] Gouveia, L. *Microalgae as a Feedstock for Biofuels*. Springer Berlin Heidelberg, 2011.

[95] Halim, R., Danquah, M. K., and Webley, P. A. Extraction of oil from microalgae for biodiesel production: A review. *Biotechnology Advances* **30** (2012), 709–732.

[96] Safi, C., Zebib, B., Merah, O., Pontalier, P.-Y., and Vaca-Garcia, C. Morphology, composition, production, processing and applications of Chlorella vulgaris: A review. *Renewable and Sustainable Energy Reviews* **35** (2014), 265–278.

[97] Davis, R., Aden, A., and Pienkos, P. T. Techno-economic analysis of autotrophic microalgae for fuel production. *Applied Energy* **88** (2011), 3524–3531.

[98] Granados, M., Acién, F., Gómez, C., Fernández-Sevilla, J., and Grima, E. M. Evaluation of flocculants for the recovery of freshwater microalgae. *Bioresource Technology* **118** (2012), 102–110.

[99] Araujo, G. S., Matos, L. J., Gonçalves, L. R., Fernandes, F. A., and Farias, W. R. Bioprospecting for oil producing microalgal strains: Evaluation of oil and biomass production for ten microalgal strains. *Bioresource Technology* **102** (2011), 5248–5250.

[100] Zheng, H., Gao, Z., Yin, J., Tang, X., Ji, X., and Huang, H. Harvesting of microalgae by flocculation with poly (γ-glutamic acid). *Bioresource Technology* **112** (2012), 212–220.

[101] Sotoft, L. F., Rong, B. G., Christensen, K. V., and Norddahl, B. Process simulation and economical evaluation of enzymatic biodiesel production plant. *Bioresource Technology* **101** (2010), 5266–5274.

[102] Yunus, N. A., Gernaey, K. V., Woodley, J. M., and Gani, R. A systematic methodology for design of tailor-made blended products. *Computers and Chemical Engineering* **66** (2014), 201–213.

[103] Wooley, R., Ruth, M., Sheehan, J., Ibsen, K., Majdeski, H., and Galvez, A. *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis Current and Futuristic Scenarios*. Tech. rep. National Renewable Energy Laboratory (NREL), 1999.

[104] Searcy, E., Flynn, P., Ghafoori, E., and Kumar, A. The relative cost of biomass energy transport. *Applied Biochemistry and Biotechnology* **137-140** (2007), 639–652.

[105] Oberdieck, R., Diangelakis, N. A., Papathanasiou, M. M., Nascu, I., and Pistikopoulos, E. N. POP – Parametric Optimization Toolbox. *Industrial & Engineering Chemistry Research* **55** (2016), 8979–8991.

[106] Giuliano, A., Cerulli, R., Poletto, M., Raiconi, G., and Barletta, D. Process pathways optimization for a lignocellulosic biorefinery producing levulinic acid, succinic acid, and ethanol. *Industrial & Engineering Chemistry Research* **55** (2016), 10699–10717.

[107] Martín, M. and Grossmann, I. E. Optimal integration of algae-switchgrass facility for the production of methanol and biodiesel. *ACS Sustainable Chemistry & Engineering* **4** (2016), 5651–5658.

[108] Yuan, Z. and Eden, M. R. Superstructure optimization of integrated fast pyrolysis-gasification for production of liquid fuels and propylene. *AIChE Journal* **62** (2016), 3155–3176.

[109] Mountraki, A., Tsakalova, M., Panteli, A., Papoutsi, A. I., and Kokossis, A. C. Integrated waste management in multiproduct biorefineries: Systems optimization and analysis of a real-life industrial plant. *Industrial & Engineering Chemistry Research* **55** (2016), 3478–3492.

[110] Matthews, L. R., Niziolek, A. M., Onel, O., Pinnaduwage, N., and Floudas, C. A. Biomass to liquid transportation fuels via biological and thermochemical conversion: Process synthesis and global optimization strategies. *Industrial & Engineering Chemistry Research* **55** (2016), 3205–3225.

[111] Satchatippavarn, S., Martinez-Hernandez, E., Hang, M. Y. L. P., Leach, M., and Yang, A. Urban biorefinery for waste processing. *Chemical Engineering Research and Design* **107** (2016), 81–90.

[112] Giuliano, A., Poletto, M., and Barletta, D. Process optimization of a multi-product biorefinery: The effect of biomass seasonality. *Chemical Engineering Research and Design* **107** (2016), 236–252.

[113] Cheali, P., Posada, J. A., Gernaey, K. V., and Sin, G. Upgrading of lignocellulosic biorefinery to value-added chemicals: Sustainability and economics of bioethanol-derivatives. *Biomass and Bioenergy* **75** (2015), 282–300.

[114] Cheali, P., Vivion, A., Gernaey, K. V., and Sin, G. *Optimal design of algae biorefinery processing networks for the production of protein, ethanol and biodiesel*. Vol. 37. Elsevier, 2015, 1151–1156.

[115] Cheali, P., Quaglia, A., Gernaey, K. V., and Sin, G. Effect of market price uncertainties on the design of optimal biorefinery systems – A systematic approach. *Industrial & Engineering Chemistry Research* **53** (2014), 6021–6032.

[116] Kelloway, A. and Daoutidis, P. Process synthesis of biorefineries: Optimization of biomass conversion to fuels and chemicals. *Industrial & Engineering Chemistry Research* **53** (2014), 5261–5273.

[117] Gong, J. and You, F. Optimal design and aynthesis of algal biorefinery processes for biological carbon sequestration and utilization with zero direct greenhouse gas emissions: MINLP model and global optimization algorithm. *Industrial & Engineering Chemistry Research* **53** (2014), 1563–1579.

[118] de Faria, D., Quaglia, A., Pessoa, F. L., and Gani, R. Early stage design of a biorefinery from castor oil. *Computer Aided Chemical Engineering* **34** (2014), 489–494.

[119] Giuliano, A., Cerulli, R., Poletto, M., Raiconi, G., and Barletta, D. Optimization of a multiproduct lignocellulosic biorefinery using a MILP approximation. *Computer Aided Chemical Engineering* **33** (2014), 1423–1428.

[120] Gong, J. and You, F. Sustainable design and synthesis of algal biorefinery for biofuel production. *Computer Aided Chemical Engineering* **33** (2014), 1429–1434.

[121] Zhang, Q., Gong, J., and You, F. Optimal superstructure-based design and synthesis of hydrocarbon biorefinery via fast pyrolysis, hydrogen production and hydroprocessing pathway. *Computer Aided Chemical Engineering*. Computer Aided Chemical Engineering **33** (2014), 175–180.

[122] Tsakalova, M., Yang, A., and Kokossis, A. C. A systems approach for the holistic screening of second generation biorefinery paths for energy and bio-based products. *Computer Aided Chemical Engineering* **33** (2014), 205–210.

[123] Slegers, P., Koetzier, B., Fasaei, F., Wijffels, R., Straten, G. van, and Boxtel, A. van. A model-based combinatorial optimisation approach for energy-efficient processing of microalgae. *Algal Research* **5** (2014), 140–157.

$\sqrt{17}$
$\Omega \int \delta e^{i\pi} = -1$
{2.7182818284}ϑφέρτυθιοπσδφγηξκλ
χ^2
$\sum$
λ

Department of Chemical and Biochemical Engineering
Technical University of Denmark
Søltofts Plads, Building 229
2800 Kgs. Lyngby
Denmark

Phone: +45 45 25 28 00
Web: www.kt.dtu.dk